Part I

Spin Zero

1

Attempts at relativistic quantum mechanics

Prerequisite: none

In order to combine quantum mechanics and relativity, we must first understand what we mean by "quantum mechanics" and "relativity". Let us begin with quantum mechanics.

Somewhere in most textbooks on the subject, one can find a list of the "axioms of quantum mechanics". These include statements along the lines of:

> *The state of the system is represented by a vector in Hilbert space.*
>
> *Observables are represented by hermitian operators.*
>
> *The measurement of an observable yields one of its eigenvalues as the result.*

and so on. We do not need to review these closely here. The axiom we need to focus on is the one that says that the time evolution of the state of the system is governed by the Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}|\psi,t\rangle = H|\psi,t\rangle \;, \tag{1.1}$$

where H is the hamiltonian operator, representing the total energy.

Let us consider a very simple system: a spinless, nonrelativistic particle with no forces acting on it. In this case, the hamiltonian is

$$H = \frac{1}{2m}\mathbf{P}^2 \;, \tag{1.2}$$

where m is the particle's mass, and $\mathbf{P}$ is the momentum operator. In the position basis, eq. (1.1) becomes

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{x},t) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{x},t) \;, \tag{1.3}$$

where $\psi(\mathbf{x},t) = \langle \mathbf{x}|\psi,t\rangle$ is the position-space wave function. We would like to generalize this to relativistic motion.

The obvious way to proceed is to take

$$H = +\sqrt{\mathbf{P}^2c^2 + m^2c^4}\,, \tag{1.4}$$

which yields the correct relativistic energy-momentum relation. If we formally expand this hamiltonian in inverse powers of the speed of light c, we get

$$H = mc^2 + \frac{1}{2m}\mathbf{P}^2 + \ldots\,. \tag{1.5}$$

This is simply a constant (the rest energy), plus the usual nonrelativistic hamiltonian, eq. (1.2), plus higher-order corrections. With the hamiltonian given by eq. (1.4), the Schrödinger equation becomes

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{x},t) = +\sqrt{-\hbar^2c^2\nabla^2 + m^2c^4}\;\psi(\mathbf{x},t)\,. \tag{1.6}$$

Unfortunately, this equation presents us with a number of difficulties. One is that it apparently treats space and time on a different footing: the time derivative appears only on the left, outside the square root, and the space derivatives appear only on the right, under the square root. This asymmetry between space and time is not what we would expect of a relativistic theory. Furthermore, if we expand the square root in powers of ∇^2, we get an infinite number of spatial derivatives acting on $\psi(\mathbf{x},t)$; this implies that eq. (1.6) is not local in space.

We can alleviate these problems by squaring the differential operators on each side of eq. (1.6) before applying them to the wave function. Then we get

$$-\hbar^2\frac{\partial^2}{\partial t^2}\psi(\mathbf{x},t) = \left(-\hbar^2c^2\nabla^2 + m^2c^4\right)\psi(\mathbf{x},t)\,. \tag{1.7}$$

This is the *Klein–Gordon equation*, and it looks a lot nicer than eq. (1.6). It is second-order in both space and time derivatives, and they appear in a symmetric fashion.

To better understand the Klein–Gordon equation, let us consider in more detail what we mean by "relativity". Special relativity tells us that physics looks the same in all inertial frames. To explain what this means, we first suppose that a certain spacetime coordinate system $(ct,\mathbf{x})$ represents (by fiat) an inertial frame. Let us define $x^0 = ct$, and write x^μ, where $\mu = 0,1,2,3$, in place of $(ct,\mathbf{x})$. It is also convenient (for reasons not at all obvious at this point) to define $x_0 = -x^0$ and $x_i = x^i$, where $i = 1,2,3$. This can be

expressed more elegantly if we first introduce the *Minkowski metric*,

$$g_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & +1 & & \\ & & +1 & \\ & & & +1 \end{pmatrix}, \tag{1.8}$$

where blank entries are zero. We then have $x_\mu = g_{\mu\nu}x^\nu$, where a repeated index is summed.

To invert this formula, we introduce the inverse of g, which is confusingly also called g, except with both indices up:

$$g^{\mu\nu} = \begin{pmatrix} -1 & & & \\ & +1 & & \\ & & +1 & \\ & & & +1 \end{pmatrix}. \tag{1.9}$$

We then have $g^{\mu\nu}g_{\nu\rho} = \delta^\mu{}_\rho$, where $\delta^\mu{}_\rho$ is the Kronecker delta (equal to one if its two indices take on the same value, zero otherwise). Now we can also write $x^\mu = g^{\mu\nu}x_\nu$.

It is a general rule that any pair of repeated (and therefore summed) indices must consist of one superscript and one subscript; these indices are said to be *contracted*. Also, any unrepeated (and therefore unsummed) indices must match (in both name and height) on the left- and right-hand sides of any valid equation.

Now we are ready to specify what we mean by an inertial frame. If the coordinates x^μ represent an inertial frame (which they do, by assumption), then so do any other coordinates $\bar{x}^\mu$ that are related by

$$\bar{x}^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu \,, \tag{1.10}$$

where $\Lambda^\mu{}_\nu$ is a *Lorentz transformation matrix* and a^μ is a *translation vector*. Both $\Lambda^\mu{}_\nu$ and a^μ are constant (that is, independent of x^μ). Furthermore, $\Lambda^\mu{}_\nu$ must obey

$$g_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma = g_{\rho\sigma} \,. \tag{1.11}$$

Eq. (1.11) ensures that the *interval* between two different spacetime points that are labeled by x^μ and x'^μ in one inertial frame, and by $\bar{x}^\mu$ and $\bar{x}'^\mu$ in another, is the same. This interval is defined to be

$$\begin{aligned}(x - x')^2 &\equiv g_{\mu\nu}(x - x')^\mu(x - x')^\nu \\ &= (\mathbf{x} - \mathbf{x}')^2 - c^2(t - t')^2 \,.\end{aligned} \tag{1.12}$$

In the other frame, we have

$$\begin{aligned}(\bar{x}-\bar{x}')^2 &= g_{\mu\nu}(\bar{x}-\bar{x}')^\mu(\bar{x}-\bar{x}')^\nu \\ &= g_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma(x-x')^\rho(x-x')^\sigma \\ &= g_{\rho\sigma}(x-x')^\rho(x-x')^\sigma \\ &= (x-x')^2\,,\end{aligned} \tag{1.13}$$

as desired.

When we say that *physics looks the same*, we mean that two observers (Alice and Bob, say) using two different sets of coordinates (representing two different inertial frames) should agree on the predicted results of all possible experiments. In the case of quantum mechanics, the simplest possibility is for Alice and Bob to agree on the value of the wave function at a particular spacetime point, a point that is called x by Alice and $\bar{x}$ by Bob. Thus if Alice's predicted wave function is $\psi(x)$, and Bob's is $\bar{\psi}(\bar{x})$, then we should have $\psi(x)=\bar{\psi}(\bar{x})$. Furthermore, in order to maintain $\psi(x)=\bar{\psi}(\bar{x})$ throughout spacetime, $\psi(x)$ and $\bar{\psi}(\bar{x})$ should obey identical equations of motion. Thus a candidate wave equation should take the same form in any inertial frame.

Let us see if this is true of the Klein–Gordon equation. We first introduce some useful notation for spacetime derivatives:

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(+\frac{1}{c}\frac{\partial}{\partial t},\nabla\right), \tag{1.14}$$

$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu} = \left(-\frac{1}{c}\frac{\partial}{\partial t},\nabla\right). \tag{1.15}$$

Note that

$$\partial^\mu x^\nu = g^{\mu\nu}\,, \tag{1.16}$$

so that our matching-index-height rule is satisfied.

If $\bar{x}$ and x are related by eq. (1.10), then $\bar{\partial}$ and ∂ are related by

$$\bar{\partial}^\mu = \Lambda^\mu{}_\nu\partial^\nu\,. \tag{1.17}$$

To check this, we note that

$$\bar{\partial}^\rho\bar{x}^\sigma = (\Lambda^\rho{}_\mu\partial^\mu)(\Lambda^\sigma{}_\nu x^\nu + a^\sigma) = \Lambda^\rho{}_\mu\Lambda^\sigma{}_\nu(\partial^\mu x^\nu) = \Lambda^\rho{}_\mu\Lambda^\sigma{}_\nu g^{\mu\nu} = g^{\rho\sigma}\,, \tag{1.18}$$

as expected. The last equality in eq. (1.18) is another form of eq. (1.11); see section 2.

We can now write eq. (1.7) as

$$-\hbar^2c^2\partial_0^2\psi(x) = (-\hbar^2c^2\nabla^2 + m^2c^4)\psi(x)\ . \tag{1.19}$$

After rearranging and identifying $\partial^2 \equiv \partial^\mu\partial_\mu = -\partial_0^2 + \nabla^2$, we have

$$(-\partial^2 + m^2c^2/\hbar^2)\psi(x) = 0\ . \tag{1.20}$$

This is Alice's form of the equation. Bob would write

$$(-\bar\partial^2 + m^2c^2/\hbar^2)\bar\psi(\bar x) = 0\ . \tag{1.21}$$

Is Bob's equation equivalent to Alice's equation? To see that it is, we set $\bar\psi(\bar x) = \psi(x)$, and note that

$$\bar\partial^2 = g_{\mu\nu}\bar\partial^\mu\bar\partial^\nu = g_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma\partial^\rho\partial^\sigma = \partial^2\ . \tag{1.22}$$

Thus, eq. (1.21) is indeed equivalent to eq. (1.20). The Klein–Gordon equation is therefore manifestly consistent with relativity: it takes the same form in every inertial frame.

This is the good news. The bad news is that the Klein–Gordon equation violates one of the axioms of quantum mechanics: eq. (1.1), the Schrödinger equation in its abstract form. The abstract Schrödinger equation has the fundamental property of being first order in the time derivative, whereas the Klein–Gordon equation is second order. This may not seem too important, but in fact it has drastic consequences. One of these is that the norm of a state,

$$\langle\psi,t|\psi,t\rangle = \int d^3x\,\langle\psi,t|\mathbf{x}\rangle\langle\mathbf{x}|\psi,t\rangle = \int d^3x\,\psi^*(x)\psi(x), \tag{1.23}$$

is not in general time independent. Thus probability is not conserved. The Klein–Gordon equation obeys relativity, but not quantum mechanics.

Dirac attempted to solve this problem (for spin-one-half particles) by introducing an extra discrete label on the wave function, to account for spin: $\psi_a(x)$, $a = 1, 2$. He then tried a Schrödinger equation of the form

$$i\hbar\frac{\partial}{\partial t}\psi_a(x) = \Bigl(-i\hbar c(\alpha^j)_{ab}\partial_j + mc^2(\beta)_{ab}\Bigr)\psi_b(x)\ , \tag{1.24}$$

where all repeated indices are summed, and α^j and β are matrices in spin-space. This equation, the *Dirac equation*, is consistent with the abstract Schrödinger equation. The state $|\psi,a,t\rangle$ carries a spin label a, and the hamiltonian is

$$H_{ab} = cP_j(\alpha^j)_{ab} + mc^2(\beta)_{ab}\ , \tag{1.25}$$

where P_j is a component of the momentum operator.

Since the Dirac equation is linear in both time and space derivatives, it has a chance to be consistent with relativity. Note that squaring the hamiltonian yields

$$(H^2)_{ab} = c^2 P_j P_k (\alpha^j \alpha^k)_{ab} + mc^3 P_j (\alpha^j \beta + \beta \alpha^j)_{ab} + (mc^2)^2 (\beta^2)_{ab} \ . \quad (1.26)$$

Since $P_j P_k$ is symmetric on exchange of j and k, we can replace $\alpha^j \alpha^k$ by its symmetric part, $\frac{1}{2}\{\alpha^j, \alpha^k\}$, where $\{A, B\} = AB + BA$ is the anticommutator. Then, if we choose matrices such that

$$\{\alpha^j, \alpha^k\}_{ab} = 2\delta^{jk}\delta_{ab} \ , \quad \{\alpha^j, \beta\}_{ab} = 0 \ , \quad (\beta^2)_{ab} = \delta_{ab} \ , \quad (1.27)$$

we will get

$$(H^2)_{ab} = (\mathbf{P}^2 c^2 + m^2 c^4)\delta_{ab} \ . \quad (1.28)$$

Thus, the eigenstates of H^2 are momentum eigenstates, with H^2 eigenvalue $\mathbf{p}^2 c^2 + m^2 c^4$. This is, of course, the correct relativistic energy-momentum relation. While it is outside the scope of this section to demonstrate it, it turns out that the Dirac equation is fully consistent with relativity, provided the Dirac matrices obey eq. (1.27). So we have apparently succeeded in constructing a quantum mechanical, relativistic theory!

There are, however, some problems. We would like the Dirac matrices to be 2×2, in order to account for electron spin. However, they must in fact be larger. To see this, note that the 2×2 Pauli matrices obey $\{\sigma^i, \sigma^j\} = 2\delta^{ij}$, and are thus candidates for the Dirac α^i matrices. However, there is no fourth matrix that anticommutes with these three (easily proven by writing down the most general 2×2 matrix and working out the three anticommutators explicitly). Also, we can show that the Dirac matrices must be even dimensional; see problem 1.1. Thus their minimum size is 4×4, and it remains for us to interpret the two extra possible "spin" states.

However, these extra states cause a more severe problem than a mere overcounting. Acting on a momentum eigenstate, H becomes the matrix $c\,\boldsymbol{\alpha}\cdot\mathbf{p} + mc^2\beta$. In problem 1.1, we find that the trace of this matrix is zero. Thus the four eigenvalues must be $+E(\mathbf{p})$, $+E(\mathbf{p})$, $-E(\mathbf{p})$, $-E(\mathbf{p})$, where $E(\mathbf{p}) = +(\mathbf{p}^2 c^2 + m^2 c^4)^{1/2}$. The negative eigenvalues are the problem: they indicate that there is no ground state. In a more elaborate theory that included interactions with photons, there seems to be no reason why a positive energy electron could not emit a photon and drop down into a negative energy state. This downward cascade could continue forever. (The same problem also arises in attempts to interpret the Klein–Gordon equation as a modified form of quantum mechanics.)

Dirac made a wildly brilliant attempt to fix this problem of negative energy states. His solution is based on an empirical fact about electrons: they obey the Pauli exclusion principle. It is impossible to put more than one of them in the same quantum state. What if, Dirac speculated, all the negative energy states were *already occupied*? In this case, a positive energy electron could not drop into one of these states, by Pauli exclusion.

Many questions immediately arise. Why do we not see the negative electric charge of this *Dirac sea* of electrons? Dirac's answer: because we are used to it. (More precisely, the physical effects of a uniform charge density depend on the boundary conditions at infinity that we impose on Maxwell's equations, and there is a choice that renders such a uniform charge density invisible.) However, Dirac noted, if one of these negative energy electrons were excited into a positive energy state (by, say, a sufficiently energetic photon), it would leave behind a *hole* in the sea of negative energy electrons. This hole would appear to have positive charge, and positive energy. Dirac therefore predicted (in 1927) the existence of the *positron*, a particle with the same mass as the electron, but opposite charge. The positron was found experimentally five years later.

However, we have now jumped from an attempt at a quantum description of a *single* relativistic particle to a theory that apparently requires an *infinite* number of particles. Even if we accept this, we still have not solved the problem of how to describe particles like photons or pions or alpha nuclei that do *not* obey Pauli exclusion.

At this point, it is worthwhile to stop and reflect on why it has proven to be so hard to find an acceptable relativistic wave equation for a single quantum particle. Perhaps there is something wrong with our basic approach.

And there is. Recall the axiom of quantum mechanics that says that "Observables are represented by hermitian operators." This is not entirely true. There is one observable in quantum mechanics that is *not* represented by a hermitian operator: time. Time enters into quantum mechanics only when we announce that the "state of the system" depends on an extra parameter t. This parameter is not the eigenvalue of any operator. This is in sharp contrast to the particle's position $\mathbf{x}$, which *is* the eigenvalue of an operator. Thus, space and time are treated very differently, a fact that is obscured by writing the Schrödinger equation in terms of the position-space wave function $\psi(\mathbf{x},t)$. Since space and time are treated asymmetrically, it is not surprising that we are having trouble incorporating a symmetry that mixes them up.

So, what are we to do?

In principle, the problem could be an intractable one: it might be *impossible* to combine quantum mechanics and relativity. In this case, there would have to be some meta-theory, one that reduces in the nonrelativistic limit to quantum mechanics, and in the classical limit to relativistic particle dynamics, but is actually neither.

This, however, turns out not to be the case. We can solve our problem, but we must put space and time on an equal footing at the outset. There are two ways to do this. One is to demote position from its status as an operator, and render it as an extra label, like time. The other is to promote time to an operator.

Let us discuss the second option first. If time becomes an operator, what do we use as the time parameter in the Schrödinger equation? Happily, in relativistic theories, there is more than one notion of time. We can use the *proper time* τ of the particle (the time measured by a clock that moves with it) as the time parameter. The coordinate time T (the time measured by a stationary clock in an inertial frame) is then promoted to an operator. In the Heisenberg picture (where the state of the system is fixed, but the operators are functions of time that obey the classical equations of motion), we would have operators $X^\mu(\tau)$, where $X^0 = T$. Relativistic quantum mechanics can indeed be developed along these lines, but it is surprisingly complicated to do so. (The many times are the problem; any monotonic function of τ is just as good a candidate as τ itself for the proper time, and this infinite redundancy of descriptions must be understood and accounted for.)

One of the advantages of considering different formalisms is that they may suggest different directions for generalizations. For example, once we have $X^\mu(\tau)$, why not consider adding some more parameters? Then we would have, for example, $X^\mu(\sigma,\tau)$. Classically, this would give us a continuous family of worldlines, what we might call a *worldsheet*, and so $X^\mu(\sigma,\tau)$ would describe a propagating *string*. This is indeed the starting point for string theory.

Thus, promoting time to an operator is a viable option, but is complicated in practice. Let us then turn to the other option, demoting position to a label. The first question is, label on what? The answer is, *on operators*. Thus, consider assigning an operator to each point $\mathbf{x}$ in space; call these operators $\varphi(\mathbf{x})$. A set of operators like this is called a *quantum field*. In the Heisenberg picture, the operators are also time dependent:

$$\varphi(\mathbf{x},t) = e^{iHt/\hbar}\varphi(\mathbf{x},0)e^{-iHt/\hbar}\,. \tag{1.29}$$

Thus, both position and (in the Heisenberg picture) time are now labels on operators; neither is itself the eigenvalue of an operator.