

Relativistic equations

1.1 Introduction

As we know, in single particle, non-relativistic quantum mechanics, we start with the Hamiltonian description of the corresponding classical, non-relativistic physical system and promote each of the observables to a Hermitian operator. The time evolution of the quantum mechanical system, in this case, is given by the time dependent Schrödinger equation which has the form

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi. \quad (1.1)$$

Here $\psi(\mathbf{x}, t)$ represents the wave function of the system which corresponds to the probability amplitude for finding the particle at the coordinate \mathbf{x} at a given time t and the Hamiltonian, H , has the generic form

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x}), \quad (1.2)$$

with \mathbf{p} denoting the momentum of the particle and $V(\mathbf{x})$ representing the potential through which the particle moves. (Throughout the book we will use a bold symbol to represent a three dimensional quantity.)

This formalism is clearly non-relativistic (non-covariant) which can be easily seen by noting that, even for a free particle, the dynamical equation (1.1) takes the form

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{\mathbf{p}^2}{2m} \psi. \quad (1.3)$$

In the coordinate basis, the momentum operator has the form

$$\mathbf{p} \rightarrow -i\hbar \nabla, \quad (1.4)$$

so that the time dependent Schrödinger equation, in this case, takes the form

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi. \quad (1.5)$$

This equation is linear in the time derivative while it is quadratic in the space derivatives. Therefore, space and time are not treated on an equal footing in this case and, consequently, the equation cannot have the same form (covariant) in different Lorentz frames. A relativistic equation, on the other hand, must treat space and time coordinates on an equal footing and remain form invariant in all inertial frames (Lorentz frames). Let us also recall that, even for a simple fundamental system such as the Hydrogen atom, the ground state electron is fairly relativistic ($\frac{v}{c}$, for the ground state electron is of the order of the fine structure constant). Consequently, there is a need to generalize the quantum mechanical description to relativistic systems. In this chapter, we will study how we can systematically develop a quantum mechanical description of a single relativistic particle and the difficulties associated with such a description.

1.2 Notations

Before proceeding any further, let us fix our notations. We note that in the three dimensional Euclidean space, which we are all familiar with, a vector is labelled uniquely by its three components. (We denote three dimensional vectors in boldface.) Thus,

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, x_3), \\ \mathbf{J} &= (J_1, J_2, J_3), \\ \mathbf{A} &= (A_1, A_2, A_3), \end{aligned} \quad (1.6)$$

where \mathbf{x} and \mathbf{J} represent respectively the position and the angular momentum vectors while \mathbf{A} stands for any arbitrary vector. In such a space, as we know, the scalar product of any two arbitrary vectors is defined to be

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i = \delta_{ij} A_i B_j = \delta^{ij} A_i B_j, \quad (1.7)$$

where repeated indices are assumed to be summed. The scalar product of two vectors is invariant under rotations of the three dimensional space which is the maximal symmetry group of the Euclidean space that leaves the origin invariant. This also allows us to define the length of a vector simply as

$$\mathbf{A}^2 = \mathbf{A} \cdot \mathbf{A} = A_i A_i = \delta_{ij} A_i A_j = \delta^{ij} A_i A_j. \quad (1.8)$$

The Kronecker delta, δ_{ij} , in this case, represents the metric of the Euclidean space and is trivial (in the sense that all the nonzero components are simply unity). Consequently, it does not matter whether we write the indices “up” or “down”. Let us note from the definition of the length of a vector in Euclidean space that, for any vector, it is necessarily positive definite, namely,

$$\begin{aligned} \mathbf{A}^2 &\geq 0, \\ \mathbf{A}^2 &= 0, \quad \text{if and only if } \mathbf{A} = 0. \end{aligned} \quad (1.9)$$

When we treat space and time on an equal footing and enlarge our three dimensional Euclidean manifold to the four dimensional space-time manifold, we can again define vectors in this manifold. However, these would now consist of four components. Namely, any point in this manifold will be specified uniquely by four coordinates and, consequently, any vector would also have four components. However, unlike the case of the Euclidean space, there are now two distinct four vectors that we can define on this manifold, namely, ($\mu = 0, 1, 2, 3$ and we are being a little sloppy in representing the four vector by what may seem like its component)

$$\begin{aligned}x^\mu &= (ct, \mathbf{x}), \\x_\mu &= (ct, -\mathbf{x}).\end{aligned}\tag{1.10}$$

Here c represents the speed of light (necessary to give the same dimension to all the components) and we note that the two four vectors simply represent the two distinct possible ways space and time components can be embedded into the four vector. On a more fundamental level, the two four vectors have distinct transformation properties under Lorentz transformations (in fact, one transforms inversely with respect to the other) and are known respectively as contravariant and covariant vectors.

The contravariant and the covariant vectors are related to each other through the metric tensor of the four dimensional manifold, commonly known as the Minkowski space, namely,

$$\begin{aligned}x_\mu &= \eta_{\mu\nu}x^\nu, \\x^\mu &= \eta^{\mu\nu}x_\nu.\end{aligned}\tag{1.11}$$

From the forms of the contravariant and the covariant vectors in (1.10) as well as using (1.11), we can immediately read out the metric tensors for the four dimensional Minkowski space which are diagonal with the signature $(+, -, -, -)$. Namely, we can write them in the matrix form as

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},\tag{1.12}$$

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.\tag{1.13}$$

The contravariant metric tensor, $\eta^{\mu\nu}$, and the covariant metric tensor, $\eta_{\mu\nu}$, are inverses of each other, since they satisfy

$$\eta^{\mu\lambda}\eta_{\lambda\nu} = \delta^{\mu}_{\nu}. \quad (1.14)$$

Furthermore, each is symmetric as they are expected to be, namely,

$$\eta^{\mu\nu} = \eta^{\nu\mu}, \quad \eta_{\mu\nu} = \eta_{\nu\mu}. \quad (1.15)$$

This particular choice of the metric is conventionally known as the Bjorken-Drell metric and this is what we will be using throughout these lectures. Different authors, however, use different metric conventions and you should be careful in reading the literature. (As is clear from the above discussion, the nonuniqueness in the choice of the metric tensors reflects the nonuniqueness of the embedding of space and time components into a four vector. Physical results, however, are independent of the choice of a metric.)

Given two arbitrary four vectors

$$\begin{aligned} A^\mu &= (A^0, \mathbf{A}), \\ B^\mu &= (B^0, \mathbf{B}), \end{aligned} \quad (1.16)$$

we can define an invariant scalar product of the two vectors as

$$\begin{aligned} A \cdot B &= A^\mu B_\mu = A_\mu B^\mu \\ &= \eta^{\mu\nu} A_\mu B_\nu = \eta_{\mu\nu} A^\mu B^\nu \\ &= A^0 B^0 - \mathbf{A} \cdot \mathbf{B}. \end{aligned} \quad (1.17)$$

Since the contravariant and the covariant vectors transform in an inverse manner, such a product is easily seen to be invariant under Lorentz transformations. This is the generalization of the scalar product of the three dimensional Euclidean space to the four dimensional Minkowski space and is invariant under Lorentz transformations which are the analogs of rotations in Minkowski space. In fact, any product of Lorentz tensors defines a scalar if all the Lorentz indices are contracted, namely, if there is no free Lorentz index. (Two Lorentz indices are said to be contracted if a contravariant and a covariant index are summed over all possible values.)

Given this, we note that the length of a (four) vector in Minkowski space can be determined to have the form

$$A^2 = A \cdot A = \eta^{\mu\nu} A_\mu A_\nu = \eta_{\mu\nu} A^\mu A^\nu = (A^0)^2 - \mathbf{A}^2. \quad (1.18)$$

Unlike the Euclidean space, however, here we see that the length of a vector need not always be positive semi-definite. In fact, if we look at the Minkowski space itself, we find that

$$x^2 = x^\mu x_\mu = \eta_{\mu\nu} x^\mu x^\nu = c^2 t^2 - \mathbf{x}^2. \quad (1.19)$$

This is the invariant length (of any point from the origin) in this space. The invariant length between two points infinitesimally close to each other follows from this to be

$$ds^2 = c^2 d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (1.20)$$

where τ is known as the proper time.

For coordinates which satisfy (see (1.19), we will set $c = 1$ from now on for simplicity)

$$x^2 = t^2 - \mathbf{x}^2 > 0, \quad (1.21)$$

we say that the region of space-time is time-like for obvious reasons. On the other hand, for coordinates which satisfy

$$x^2 = t^2 - \mathbf{x}^2 < 0, \quad (1.22)$$

the region of space-time is known as space-like. The boundary of the two regions, namely, the region for which

$$x^2 = t^2 - \mathbf{x}^2 = 0, \quad (1.23)$$

defines trajectories for light-like particles and is, consequently, known as the light-like region. (Light-like vectors, for which the invariant length vanishes, are nontrivial unlike the case of the Euclidean space.)

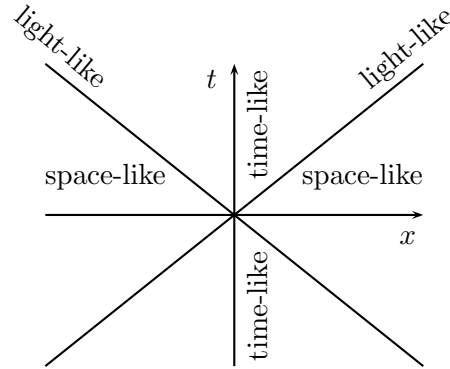


Figure 1.1: Different invariant regions of Minkowski space.

Thus, we see that, unlike the Euclidean space, the Minkowski space-time manifold separates into four invariant wedges (regions which do not mix under Lorentz transformations), which in a two dimensional projection has the form shown in Fig. 1.1. The different invariant wedges are known as

$$\begin{aligned}
 t > 0, \quad x^2 \geq 0 &: \text{ future light cone,} \\
 t < 0, \quad x^2 \geq 0 &: \text{ past light cone,} \\
 x^2 < 0 &: \text{ space - like.}
 \end{aligned} \tag{1.24}$$

All physical processes are assumed to take place in the future light cone or the forward light cone defined by

$$t > 0 \quad \text{and} \quad x^2 \geq 0. \tag{1.25}$$

Given the contravariant and the covariant coordinates, we can define the contragradient and the cgradient respectively as ($c = 1$)

$$\begin{aligned}\partial^\mu &= \frac{\partial}{\partial x_\mu} = \left(\frac{\partial}{\partial t}, -\nabla \right), \\ \partial_\mu &= \frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial t}, \nabla \right).\end{aligned}\tag{1.26}$$

From these, we can construct the Lorentz invariant quadratic operator

$$\square = \partial^2 = \partial^\mu \partial_\mu = \frac{\partial^2}{\partial t^2} - \nabla^2,\tag{1.27}$$

which is known as the D'Alembertian. It is the generalization of the Laplacian to the four dimensional Minkowski space.

Let us note next that energy and momentum also define four vectors in this case. (Namely, they transform like four vectors under Lorentz transformations.) Thus, we can write (remember that $c = 1$, otherwise, we have to write $\frac{E}{c}$)

$$\begin{aligned}p^\mu &= (E, \mathbf{p}), \\ p_\mu &= (E, -\mathbf{p}).\end{aligned}\tag{1.28}$$

Given the energy-momentum four vectors, we can construct the Lorentz scalar

$$p^2 = p^\mu p_\mu = E^2 - \mathbf{p}^2.\tag{1.29}$$

The Einstein relation for a free particle (remember $c = 1$)

$$E^2 = \mathbf{p}^2 + m^2,\tag{1.30}$$

where m represents the rest mass of the particle, can now be seen as the Lorentz invariant condition

$$p^2 = E^2 - \mathbf{p}^2 = m^2.\tag{1.31}$$

In other words, in this space, the energy and the momentum of a free particle must lie on a hyperbola satisfying the above relation.

We already know that the coordinate representation of the energy and the momentum operators takes the forms

$$\begin{aligned} E &\rightarrow i\hbar\frac{\partial}{\partial t}, \\ \mathbf{p} &\rightarrow -i\hbar\nabla. \end{aligned} \quad (1.32)$$

We can combine these to write the coordinate representation for the energy-momentum four vector operator as

$$\begin{aligned} p^\mu &= i\hbar\partial^\mu = i\hbar\frac{\partial}{\partial x_\mu} = \left(i\hbar\frac{\partial}{\partial t}, -i\hbar\nabla \right), \\ p_\mu &= i\hbar\partial_\mu = i\hbar\frac{\partial}{\partial x^\mu} = \left(i\hbar\frac{\partial}{\partial t}, i\hbar\nabla \right). \end{aligned} \quad (1.33)$$

Finally, let us note that in the four dimensional space-time, we can construct two totally antisymmetric fourth rank tensors $\epsilon^{\mu\nu\lambda\rho}, \epsilon_{\mu\nu\lambda\rho}$, the four dimensional contravariant and covariant Levi-Civita tensors respectively. We will choose the normalization $\epsilon^{0123} = 1 = -\epsilon_{0123}$ so that

$$\epsilon^{0ijk} = \epsilon_{ijk} = -\epsilon_{0ijk}, \quad (1.34)$$

where ϵ_{ijk} denotes the three dimensional Levi-Civita tensor with $\epsilon_{123} = 1$. An anti-symmetric tensor such as $\epsilon^i{}_{jk}$ is then understood to denote

$$\epsilon^i{}_{jk} = \eta^{i\ell}\epsilon_{\ell jk}, \quad (1.35)$$

and so on. This completes the review of all the essential basic notation that we will be using in this book. We will introduce new notations as they arise in the context of our discussions.

1.3 Klein-Gordon equation

With all these basics, we are now ready to write down the simplest of the relativistic equations. Note that in the case of a non-relativistic particle, we start with the non-relativistic energy-momentum relation

$$E = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x}), \quad (1.36)$$

and promote the dynamical variables (observables) to Hermitian operators to obtain the time-dependent Schrödinger equation (see (1.1))

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}) \right) \psi. \quad (1.37)$$

Let us consider the simplest of relativistic systems, namely, a relativistic free particle of mass m . In this case, we have seen that the energy-momentum relation is none other than the Einstein relation (1.30), namely,

$$\begin{aligned} E^2 &= \mathbf{p}^2 + m^2, \\ \text{or, } \quad E^2 - \mathbf{p}^2 &= p^\mu p_\mu = m^2. \end{aligned} \quad (1.38)$$

Thus, as before, promoting these to operators, we obtain the simplest relativistic quantum mechanical equation to be (see (1.33))

$$\begin{aligned} p^\mu p_\mu \phi &= m^2 \phi, \\ \text{or, } \quad (i\hbar \partial^\mu)(i\hbar \partial_\mu) \phi &= m^2 \phi, \\ \text{or, } \quad -\hbar^2 \square \phi &= m^2 \phi. \end{aligned} \quad (1.39)$$

Setting $\hbar = 1$ from now on for simplicity, the equation above takes the form

$$(\square + m^2)\phi = 0. \quad (1.40)$$

Since the operator in the parenthesis is a Lorentz scalar and since we assume the quantum mechanical wave function, $\phi(\mathbf{x}, t)$, to be a scalar function, this equation is invariant under Lorentz transformations.

This equation, (1.40), is known as the Klein-Gordon equation and, for $m = 0$, or when the rest mass vanishes, it reduces to the wave equation (recall Maxwell's equations). Like the wave equation, the Klein-Gordon equation also has plane wave solutions which are characteristic of free particle solutions. In fact, the functions

$$e^{\mp ik \cdot x} = e^{\mp ik_\mu x^\mu} = e^{\mp ik^\mu x_\mu} = e^{\mp i(k_0 t - \mathbf{k} \cdot \mathbf{x})}, \quad (1.41)$$

with $k^\mu = (k^0, \mathbf{k})$ are eigenfunctions of the energy-momentum operator, namely, using (1.33) (remember that $\hbar = 1$) we obtain

$$p^\mu e^{\mp ik \cdot x} = i \partial^\mu e^{\mp ik \cdot x} = i \frac{\partial}{\partial x_\mu} e^{\mp ik \cdot x} = \pm k^\mu e^{\mp ik \cdot x}, \quad (1.42)$$

so that $\pm k^\mu$ are the eigenvalues of the energy-momentum operator. (In fact, the eigenvalues should be $\pm \hbar k^\mu$, but we have set $\hbar = 1$.) This shows that the plane waves define a solution of the Klein-Gordon equation provided

$$\begin{aligned} k^2 - m^2 &= (k^0)^2 - \mathbf{k}^2 - m^2 = 0, \\ \text{or, } k^0 &= \pm E = \pm \sqrt{\mathbf{k}^2 + m^2}. \end{aligned} \quad (1.43)$$

Thus, we see the first peculiarity of the Klein-Gordon equation (which is a relativistic equation), namely, that it allows for both positive and negative energy solutions. This basically arises from the fact that, for a relativistic particle (even a free one), the energy-momentum relation is given by the Einstein relation which is a quadratic relation in E , as opposed to the case of a non-relativistic particle, where the energy-momentum relation is linear in E . If we accept the Klein-Gordon equation as describing a free, relativistic, quantum mechanical particle of mass m , then, we will see shortly that the presence of the negative energy solutions would render the theory inconsistent.

To proceed further, let us note that the Klein-Gordon equation and its complex conjugate (remember that a quantum mechanical wave function is, in general, complex), namely,

$$\begin{aligned}(\square + m^2)\phi &= 0, \\(\square + m^2)\phi^* &= 0,\end{aligned}\tag{1.44}$$

would imply

$$\begin{aligned}\phi^*\square\phi - \phi\square\phi^* &= 0, \\ \text{or, } \quad \partial_\mu(\phi^*\partial^\mu\phi - \phi\partial^\mu\phi^*) &= 0, \\ \text{or, } \quad \frac{\partial}{\partial t}\left(\phi^*\frac{\partial\phi}{\partial t} - \phi\frac{\partial\phi^*}{\partial t}\right) - \nabla \cdot (\phi^*\nabla\phi - \phi\nabla\phi^*) &= 0.\end{aligned}\tag{1.45}$$

Defining the probability current density four vector as

$$J^\mu = (j^0, \mathbf{J}) = (\rho, \mathbf{J}),\tag{1.46}$$

where

$$\begin{aligned}\mathbf{J} &= \frac{1}{2im}(\phi^*\nabla\phi - \phi\nabla\phi^*), \\ \rho &= \frac{i}{2m}\left(\phi^*\frac{\partial\phi}{\partial t} - \phi\frac{\partial\phi^*}{\partial t}\right),\end{aligned}\tag{1.47}$$

we note that equation (1.45) can be written as a continuity equation for the probability current, namely,

$$\partial_\mu J^\mu = \frac{\partial\rho}{\partial t} + \nabla \cdot \mathbf{J} = 0.\tag{1.48}$$

The probability current density,

$$\mathbf{J} = \frac{1}{2im}(\phi^*\nabla\phi - \phi\nabla\phi^*),\tag{1.49}$$

of course, has the same form as in non-relativistic quantum mechanics. However, we note that the form of the probability density (which results from the requirement of covariance)

$$\rho = \frac{i}{2m} \left(\phi^* \frac{\partial \phi}{\partial t} - \phi \frac{\partial \phi^*}{\partial t} \right), \quad (1.50)$$

is quite different from that in non-relativistic quantum mechanics and it is here that the problem of the negative energy states shows up. For example, even for the simplest of solutions, namely, plane waves of the form

$$\phi(x) = e^{-ik \cdot x}, \quad (1.51)$$

we obtain

$$\rho = \frac{i}{2m} (-ik^0 - ik^0) = \frac{k^0}{m} = \pm \frac{E}{m}. \quad (1.52)$$

Since energy can take both positive and negative values, it follows that ρ cannot truly represent the probability density which, by definition, has to be positive semi-definite. It is worth noting here that this problem really arises because the Klein-Gordon equation, unlike the time dependent Schrödinger equation, is second order in time derivatives. This has the consequence that the probability density involves a first order time derivative and that is how the problem of the negative energy states enters. (Note that if the equation is second order in the space derivatives, then covariance would require that it be second order in time derivative as well. This would, in turn, lead to the difficulty with the probability density being positive semi-definite.) One can, of course, ask whether we can restrict ourselves to positive energy solutions only in order to avoid the difficulty with the interpretation of ρ . Classically, we can do this. However, quantum mechanically, we cannot arbitrarily impose this for a variety of reasons. The simplest way to see this is to note that the positive energy solutions alone do not define a complete set of states (basis) in the Hilbert space and, consequently, even if we restrict the states to be of positive energy to begin with, negative energy states may be

generated through quantum mechanical corrections. It is for these reasons that the Klein-Gordon equation was abandoned as a quantum mechanical equation for a single relativistic particle. However, as we will see later, this equation is quite meaningful as a relativistic field equation.

1.3.1 Klein paradox. Let us consider a charged scalar particle described by the Klein-Gordon equation (1.40) in an external electromagnetic field. We recall that the coupling of a charged particle to an electromagnetic field is given by the minimal coupling

$$\begin{aligned} p_\mu &\rightarrow p_\mu - eA_\mu, \\ \text{or, } \quad \partial_\mu &\rightarrow \partial_\mu + ieA_\mu, \end{aligned} \tag{1.53}$$

where we have used the coordinate representation for the momentum as in (1.33) and A_μ denotes the vector potential associated with the electromagnetic field. In this case, therefore, the scalar particle will satisfy the minimally coupled Klein-Gordon equation

$$((\partial_\mu + ieA_\mu)(\partial^\mu + ieA^\mu) + m^2)\phi(x) = 0. \tag{1.54}$$

As a result, the probability current density in (1.46) can be determined to have the form

$$J^\mu = \frac{i}{2m} \left(\phi^*(x) \overleftrightarrow{\partial}^\mu \phi(x) + 2ieA^\mu \phi^*(x)\phi(x) \right), \tag{1.55}$$

where we have defined

$$A \overleftrightarrow{\partial}^\mu B = A(\partial^\mu B) - (\partial^\mu A)B. \tag{1.56}$$

With this general description, let us consider the scattering of a charged scalar (Klein-Gordon) particle with positive energy from a constant electrostatic potential. In this case, therefore, we have

$$\mathbf{A} = 0, \quad A^0 = \Phi = \text{constant}. \tag{1.57}$$

For simplicity, let us assume the constant electrostatic potential to be of the form

$$\Phi(z) = \begin{cases} 0, & z < 0, \\ \Phi_0, & z > 0, \end{cases} \quad (1.58)$$

and we assume that the particle is incident on the potential along the z -axis as shown in Fig. 1.2.

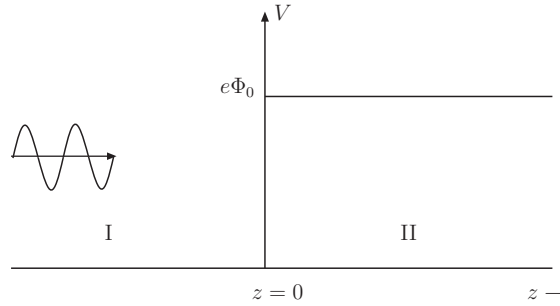


Figure 1.2: Klein-Gordon particle scattering from a constant electrostatic potential.

The dynamical equations will now be different in the two regions, $z < 0$ (region I) and $z > 0$ (region II), and have the forms

$$\begin{aligned} (\square + m^2) \phi_{\text{I}} &= 0, & z < 0, \\ \left(\square + m^2 + 2ie\Phi_0 \frac{\partial}{\partial t} - e^2\Phi_0^2 \right) \phi_{\text{II}} &= 0, & z > 0. \end{aligned} \quad (1.59)$$

In region I, there will be an incident as well as a reflected wave so that we can write

$$\phi_{\text{I}}(t, z) = e^{-iEt} (e^{ipz} + A e^{-ipz}), \quad z < 0, \quad (1.60)$$

while in region II, we only expect a transmitted wave of the form

$$\phi_{\text{II}}(t, z) = B e^{-iEt + ip'z}, \quad z > 0, \quad (1.61)$$

where A, B are related respectively to reflection and transmission coefficients. We note here that the continuity of the wave function at the boundary $z = 0$ requires that the energy be the same in the two regions.

For the wave functions in (1.60) and (1.61) to satisfy the respective equations in (1.59), we must have

$$\begin{aligned} E &= \sqrt{\mathbf{p}^2 + m^2}, \\ p' &= \pm \sqrt{(E - e\Phi_0)^2 - m^2} \\ &= \pm \sqrt{(E - e\Phi_0 + m)(E - e\Phi_0 - m)}. \end{aligned} \quad (1.62)$$

Here we have used the fact that the energy of the incident particle is positive and, therefore, the square root in the first equation in (1.62) is with a positive sign. However, the sign of the square root in the second relation remains to be fixed.

Let us note from the second relation in (1.62) that p' is real for both $E - e\Phi_0 > m$ (weak potential) and for $E - e\Phi_0 < -m$ (strong potential). However, for a potential of intermediate strength satisfying $-m < E - e\Phi_0 < m$, we note that p' is purely imaginary. Thus, the behavior of the transmitted wave depends on the strength of the potential. As a result, in this second case, we must have

$$p' = i|p'|, \quad -m < E - e\Phi_0 < m, \quad (1.63)$$

in order that the wave function is damped in region II. To determine the sign of the square root in the cases when p' is real, let us note from the second relation in (1.62) that the group velocity of the transmitted wave is given by

$$v_{\text{group}} = \frac{\partial E}{\partial p'} = \frac{p'}{E - e\Phi_0}. \quad (1.64)$$

Since we expect the transmitted wave to be travelling to the right, we determine from (1.64) that

$$\begin{aligned} p' &> 0, \quad \text{for } E - e\Phi_0 > 0, \\ p' &< 0 \quad \text{for } E - e\Phi_0 < 0. \end{aligned} \quad (1.65)$$

This, therefore, fixes the sign of the square root in the second relation in (1.62) for various cases.

Matching the wave functions in (1.60) and (1.61) and their first derivatives at the boundary $z = 0$, we determine

$$1 + A = B, \quad 1 - A = \frac{p'}{p} B, \quad (1.66)$$

so that we determine

$$A = \frac{p - p'}{p + p'}, \quad B = \frac{2p}{p + p'}. \quad (1.67)$$

Let us next determine the probability current densities associated with the different beams. From (1.55) as well as the form of the potential in (1.58) we obtain

$$\begin{aligned} J_{\text{inc}} &= \hat{\mathbf{z}} \cdot \mathbf{J}_{\text{inc}} = \frac{p}{m}, \\ J_{\text{refl}} &= -\hat{\mathbf{z}} \cdot \mathbf{J}_{\text{refl}} = \frac{p}{m} \frac{(p - p')(p - (p')^*)}{(p + p')(p + (p')^*)}, \\ J_{\text{trans}} &= \hat{\mathbf{z}} \cdot \mathbf{J}_{\text{trans}} = \frac{(p' + (p')^*)}{2m} \frac{4p^2}{(p + p')(p + (p')^*)}, \end{aligned} \quad (1.68)$$

where we have used (1.67) as well as the fact that, while p is real and positive, p' can be positive or negative or even imaginary depending on the strength of the potential (see (1.63) and (1.65)). We can now determine the reflection and the transmission coefficients simply as

$$\begin{aligned} R &= \frac{J_{\text{refl}}}{J_{\text{inc}}} = \frac{(p - p')(p - (p')^*)}{(p + p')(p + (p')^*)}, \\ T &= \frac{J_{\text{trans}}}{J_{\text{inc}}} = \frac{2p(p' + (p')^*)}{(p + p')(p + (p')^*)}. \end{aligned} \quad (1.69)$$

We see from the reflection and the transmission coefficients that

$$R + T = \frac{(p - p')(p - (p')^*) + 2p(p' + (p')^*)}{(p + p')(p + (p')^*)} = 1, \quad (1.70)$$

so that the reflection and the transmission coefficients satisfy unitarity for all strengths of the potential.

However, let us now analyze the different cases of the potential strengths individually. First, for the case, $E - e\Phi_0 > m$ (weak potential), we see that p' is real and positive and we have

$$R = \left(\frac{p - p'}{p + p'} \right)^2 < 1, \quad T = \frac{4pp'}{(p + p')^2} > 0, \quad R + T = 1, \quad (1.71)$$

which corresponds to the normal scenario in scattering. For the case of an intermediate potential strength, $-m < E - e\Phi_0 < m$, we note from (1.63) that p' is purely imaginary in this case. As a result, it follows from (1.69) that

$$R = \frac{(p - p')(p + p')}{(p + p')(p - p')} = 1, \quad T = 0, \quad R + T = 1, \quad (1.72)$$

so that the incident beam is totally reflected and there is no transmission in this case. The third case of the strong potential, $E - e\Phi_0 < -m$, is the most interesting. In this case, we note from (1.65) that p' is real, but negative. As a result, from (1.69) we have

$$R = \left(\frac{p + |p'|}{p - |p'|} \right)^2 > 1, \quad T = -\frac{4p|p'|}{(p - |p'|)^2} < 0, \quad R + T = 1. \quad (1.73)$$

Namely, even though unitarity is not violated, in this case the transmission coefficient is negative and the reflection coefficient exceeds unity. This is known as the Klein paradox and it contradicts our intuition from the one particle scatterings studied in non-relativistic quantum mechanics. On the other hand, if we go beyond the one particle description and assume that a sufficiently strong enough electrostatic potential can produce particle-antiparticle pairs, there is no paradox. For example, the antiparticles are attracted by the barrier leading to a negative charged current moving to the right which explains the negative transmission coefficient. On the other hand, the particles are reflected from the barrier and add to the totally reflected incident particles (which is already seen for intermediate strength potentials) to give a reflection coefficient that exceeds unity.

1.4 Dirac equation

As we have seen, relativistic equations seem to imply the presence of both positive as well as negative energy solutions and that quantum mechanically, we need both these solutions to describe a physical system. Furthermore, as we have seen, the Klein-Gordon equation is second order in the time derivatives and this leads to the definition of the probability density which is first order in the time derivative. Together with the negative energy solutions, this implies that the probability density can become negative which is inconsistent with the definition of a probability density. It is clear, therefore, that even if we cannot avoid the negative energy solutions, we can still possibly obtain a consistent probability density provided we have a relativistic equation which is first order in the time derivative just like the time dependent Schrödinger equation. The difference, of course, is that Lorentz invariance would require space and time to be treated on an equal footing and, therefore, such an equation, if we can find it, must be first order in both space and time derivatives. Clearly, this can be done provided we have a linear relation between energy and momentum operators. Let us recall that the Einstein relation gives

$$E^2 = \mathbf{p}^2 + m^2. \quad (1.74)$$

The positive square root of this gives

$$E = \sqrt{\mathbf{p}^2 + m^2}, \quad (1.75)$$

which is far from a linear relation.

Although the naive square root of the Einstein relation does not lead to a linear relation between the energy and the momentum variables, a matrix square root may, in fact, lead to such a relation. This is exactly what Dirac proposed. Let us, for example, write the Einstein relation as

$$\begin{aligned} E^2 - \mathbf{p}^2 &= m^2, \\ \text{or, } p^2 &= p^\mu p_\mu = m^2. \end{aligned} \quad (1.76)$$

Let us consider this as a matrix relation (namely, an $n \times n$ identity matrix multiplying both sides). Let us further assume that there exist four linearly independent $n \times n$ matrices γ^μ , $\mu = 0, 1, 2, 3$, which are space-time independent such that

$$\not{p} = \gamma^\mu p_\mu, \quad (1.77)$$

represents the matrix square root of p^2 . If this is true, then, by definition, we have

$$\begin{aligned} \not{p}\not{p} &= p^2 \mathbb{1}, \\ \text{or, } \quad \gamma^\mu p_\mu \gamma^\nu p_\nu &= p^2 \mathbb{1}, \\ \text{or, } \quad \gamma^\mu \gamma^\nu p_\mu p_\nu &= p^2 \mathbb{1}, \\ \text{or, } \quad \frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) p_\mu p_\nu &= p^2 \mathbb{1}. \end{aligned} \quad (1.78)$$

Here $\mathbb{1}$ denotes the identity matrix (in the appropriate matrix space, in this case, n dimensional) and we have used the fact that the matrices, γ^μ , are constant to move them past the momentum operators. For the relation (1.78) to be true, it is clear that the matrices, γ^μ , have to satisfy the algebra ($\mu = 0, 1, 2, 3$)

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = [\gamma^\mu, \gamma^\nu]_+ = 2\eta^{\mu\nu} \mathbb{1}. \quad (1.79)$$

Here the brackets with a subscript “+” stand for the anti-commutator of two quantities defined in (1.79) (sometimes it is also denoted by curly brackets which we will not use to avoid confusion with Poisson brackets) and this algebra is known as the Clifford algebra. We see that if we can find a set of four linearly independent constant matrices satisfying the Clifford algebra, then, we can obtain a matrix square root of p^2 which would be linear in energy and momentum.

Before going into an actual determination of such matrices, let us look at the consequences of such a possibility. In this case, the solutions of the equation (sign of the mass term is irrelevant and the mass term is multiplied by the identity matrix which we do not write explicitly)

$$\not{p}\psi = m\psi, \quad (1.80)$$

would automatically satisfy the Einstein relation. Namely,

$$\begin{aligned} \not{p}(\not{p}\psi) &= m\not{p}\psi, \\ \text{or, } p^2\psi &= m^2\psi. \end{aligned} \quad (1.81)$$

Furthermore, since the new equation is linear in the energy and momentum variables, it will, consequently, be linear in the space and time derivatives. This is, of course, what we would like for a consistent definition of the probability density. The equation (1.80) (or its coordinate representation) is known as the Dirac equation.

To determine the matrices, γ^μ , and their dimensionality, let us note that the Clifford algebra in (1.79)

$$[\gamma^\mu, \gamma^\nu]_+ = 2\eta^{\mu\nu}\mathbb{1}, \quad \mu = 0, 1, 2, 3, \quad (1.82)$$

can be written out explicitly as

$$\begin{aligned} (\gamma^0)^2 &= \mathbb{1}, \\ (\gamma^i)^2 &= -\mathbb{1}, \quad \text{for any fixed } i = 1, 2, 3, \\ \gamma^0\gamma^i + \gamma^i\gamma^0 &= 0, \\ \gamma^i\gamma^j + \gamma^j\gamma^i &= 0, \quad i \neq j. \end{aligned} \quad (1.83)$$

We can choose any one of the matrices to be diagonal and without loss of generality, let us choose

$$\gamma^0 = \begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{pmatrix}. \quad (1.84)$$

From the fact that $(\gamma^0)^2 = \mathbb{1}$, we conclude that each of the diagonal elements in γ^0 must be ± 1 , namely,

$$b_\alpha = \pm 1, \quad \alpha = 1, 2, \dots, n. \quad (1.85)$$

Let us note next that using the relations from the Clifford algebra in (1.83), for a fixed i , we obtain

$$\text{Tr } \gamma^i \gamma^0 \gamma^i = \text{Tr } \gamma^i (-\gamma^i \gamma^0) = -\text{Tr } (\gamma^i)^2 \gamma^0 = \text{Tr } \gamma^0, \quad (1.86)$$

where “Tr” denotes trace over the matrix indices. On the other hand, the cyclicity property of the trace, namely,

$$\text{Tr } ABC = \text{Tr } CAB, \quad (1.87)$$

leads to

$$\text{Tr } \gamma^i \gamma^0 \gamma^i = \text{Tr } (\gamma^i)^2 \gamma^0 = -\text{Tr } \gamma^0. \quad (1.88)$$

Thus, comparing Eqs. (1.86) and (1.88), we obtain

$$\begin{aligned} \text{Tr } \gamma^i \gamma^0 \gamma^i &= \text{Tr } \gamma^0 = -\text{Tr } \gamma^0, \\ \text{or, } \quad \text{Tr } \gamma^0 &= 0. \end{aligned} \quad (1.89)$$

For this to be true, we conclude that γ^0 must have as many diagonal elements with value $+1$ as with -1 . Consequently, the γ^μ matrices must be even dimensional.

Let us assume that $n = 2N$. The simplest nontrivial matrix structure would arise for $N = 1$ when the matrices would be two dimensional (namely, 2×2 matrices). We know that the three Pauli matrices along with the identity matrix define a complete basis for 2×2 matrices. However, as we know, they do not satisfy the Clifford algebra. Namely, if we define $\sigma^\mu = (\mathbb{1}, \boldsymbol{\sigma})$, then,

$$[\sigma^\mu, \sigma^\nu]_+ \neq 2\eta^{\mu\nu} \mathbb{1}. \quad (1.90)$$

In fact, we know that in two dimensions, there cannot exist four anti-commuting matrices.

The next choice is $N = 2$ for which the matrices will be four dimensional (4×4 matrices). In this case, we can find a set of four linearly independent, constant matrices which satisfy the Clifford algebra. A particular choice of these matrices, for example, has the form

$$\begin{aligned}\gamma^0 &= \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}, \\ \gamma^i &= \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad i = 1, 2, 3,\end{aligned}\tag{1.91}$$

where each element of the 4×4 matrices represents a 2×2 matrix and the σ_i correspond to the three Pauli matrices. This particular choice of the Dirac matrices is commonly known as the Pauli-Dirac representation.

There are, of course, other representations for the γ^μ matrices. However, the physics of Dirac equation is independent of any particular representation for the γ^μ matrices. This can be easily seen by invoking Pauli's fundamental theorem which says that if there are two sets of matrices γ^μ and γ'^μ satisfying the Clifford algebra, then, they must be related by a similarity transformation. Namely, if

$$\begin{aligned}[\gamma^\mu, \gamma^\nu]_+ &= 2\eta^{\mu\nu} \mathbb{1}, \\ [\gamma'^\mu, \gamma'^\nu]_+ &= 2\eta^{\mu\nu} \mathbb{1},\end{aligned}\tag{1.92}$$

then, there exists a constant, nonsingular matrix S such that (in fact, the similarity transformation is really a unitary transformation if we take the Hermiticity properties of the γ -matrices)

$$\gamma'^\mu = S\gamma^\mu S^{-1}.\tag{1.93}$$

Therefore, given the equation

$$(\gamma'^\mu p_\mu - m)\psi' = 0,\tag{1.94}$$

we obtain

$$\begin{aligned}
& (S\gamma^\mu S^{-1}p_\mu - m)\psi' = 0, \\
\text{or, } & S(\gamma^\mu p_\mu - m)S^{-1}\psi' = 0, \\
\text{or, } & (\gamma^\mu p_\mu - m)S^{-1}\psi' = 0, \\
\text{or, } & (\gamma^\mu p_\mu - m)\psi = 0,
\end{aligned} \tag{1.95}$$

with $\psi = S^{-1}\psi'$. (The matrix S^{-1} can be moved past the momentum operator since it is assumed to be constant.) This shows that different representations of the γ^μ matrices are equivalent and merely correspond to a change in the basis of the wave function. As we know, a change of basis does not change physics.

To obtain the Hamiltonian for the Dirac equation, let us go to the coordinate representation where the Dirac equation (1.80) takes the form (remember $\hbar = 1$)

$$\begin{aligned}
& (i\cancel{\partial} - m)\psi = (i\gamma^\mu\partial_\mu - m)\psi = 0, \\
\text{or, } & (i\gamma^0\partial_0 + i\boldsymbol{\gamma} \cdot \boldsymbol{\nabla} - m)\psi = 0.
\end{aligned} \tag{1.96}$$

Multiplying with γ^0 from the left and using the fact that $(\gamma^0)^2 = \mathbb{1}$, we obtain

$$i\frac{\partial\psi}{\partial t} = (-i\gamma^0\boldsymbol{\gamma} \cdot \boldsymbol{\nabla} + m\gamma^0)\psi. \tag{1.97}$$

Conventionally, one denotes

$$\beta = \gamma^0, \quad \boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma}. \tag{1.98}$$

In terms of these matrices, then, we can write (1.97) as

$$i\frac{\partial\psi}{\partial t} = (-i\boldsymbol{\alpha} \cdot \boldsymbol{\nabla} + m\beta)\psi = (\boldsymbol{\alpha} \cdot \mathbf{p} + \beta m)\psi. \tag{1.99}$$

This is a first order equation (in time derivative) like the Schrödinger equation and we can identify the Hamiltonian for the Dirac equation with

$$H = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m. \quad (1.100)$$

In the particular representation of the γ^μ matrices in (1.91), we note that

$$\begin{aligned} \beta &= \gamma^0 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}, \\ \boldsymbol{\alpha} &= \gamma^0 \boldsymbol{\gamma} = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}. \end{aligned} \quad (1.101)$$

We can now determine either from the definition in (1.98) and (1.79) or from the explicit representation in (1.101) that the matrices $\boldsymbol{\alpha}, \beta$ satisfy the anti-commutation relations

$$\begin{aligned} [\alpha^i, \alpha^j]_+ &= 2\delta^{ij} \mathbb{1}, \\ [\alpha^i, \beta]_+ &= 0, \end{aligned} \quad (1.102)$$

with $\beta^2 = \mathbb{1}$. We can, of course, directly check from this explicit representation that both β and $\boldsymbol{\alpha}$ are Hermitian matrices. But, independently, we also note from the form of the Hamiltonian in (1.100) that, in order for it to be Hermitian, we must have

$$\begin{aligned} \beta^\dagger &= \beta, \\ \boldsymbol{\alpha}^\dagger &= \boldsymbol{\alpha}. \end{aligned} \quad (1.103)$$

In terms of the γ^μ matrices, this translates to

$$\begin{aligned} \beta &= \gamma^0 = (\gamma^0)^\dagger = \beta^\dagger, \\ \boldsymbol{\alpha} &= \gamma^0 \boldsymbol{\gamma} = (\gamma^0 \boldsymbol{\gamma})^\dagger = \boldsymbol{\alpha}^\dagger. \end{aligned} \quad (1.104)$$

Equivalently, we can write

$$\begin{aligned}(\gamma^0)^\dagger &= \gamma^0, \\ (\gamma^i)^\dagger &= -\gamma^i.\end{aligned}\tag{1.105}$$

Namely, independent of the representation, the γ^μ matrices must satisfy the Hermiticity properties in (1.105). (With a little bit of more analysis, it can be seen that, in general, the Hermiticity properties of the γ^μ matrices are related to the choice of the metric tensor and this particular choice is associated with the Bjorken-Drell metric.) In the next chapter, we would study the plane wave solutions of the first order Dirac equation.

1.5 References

The material presented in this chapter is covered in many standard textbooks and we list below only a few of them.

1. L. I. Schiff, *Quantum Mechanics*, McGraw-Hill, New York, 1968.
2. J. D. Bjorken and S. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill, New York, 1964.
3. C. Itzykson and J-B. Zuber, *Quantum Field Theory*, McGraw-Hill, New York, 1980.
4. A. Das, *Lectures on Quantum Mechanics*, Hindustan Publishing, New Delhi, India, 2003.