Chapter 1

Relativistic quantum mechanics

1.1 Relativistic notation

When we specify an instant of time t and a point (x, y, z) in space, we are actually defining a point (or event) in space-time. We denote the coordinates of this point by (x^0, x^1, x^2, x^3), where

$$x^0 \equiv ct, \quad x^1 \equiv x, \quad x^2 \equiv y, \quad x^3 \equiv z\,. \tag{1.1}$$

The indices $0, 1, 2, 3$ label the components of a four-dimensional vector along the axes t, x, y, z respectively. We shall use the Greek letters $\mu, \nu, \ldots$ to label the space-time components of these four-vectors, so that

$$x^\mu = (x^0, \mathbf{x}) = (x^0, x^1, x^2, x^3)\,, \tag{1.2}$$

where the boldfaced quantity $\mathbf{x}$ denotes the position three-vector in ordinary space.

The space-time metric tensor is given by

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{1.3}$$

or in components

$$g_{00} = 1, \quad g_{11} = g_{22} = g_{33} = -1\,, \tag{1.4}$$

with all others zero. This metric defines Minkowski space. By contrast, the metric for Euclidean space has the same sign +1 for all its diagonal elements.

Vectors like x^μ are known as *contravariant* vectors. In non-Euclidean spaces, there is another class of vectors known as *covariant* vectors. These two classes of vectors are distinguished notationally by the placement of their indices. Contravariant indices are placed as superscripts, while covariant indices as subscripts. Thus a^μ denotes a contravariant vector. The corresponding covariant vector a_μ is obtained by using the metric tensor:

$$a_\mu = \sum_{\nu=0}^{3} g_{\mu\nu} a^\nu . \tag{1.5}$$

Using (1.3), this gives

$$a_0 = a^0, \quad a_1 = -a^1, \quad a_2 = -a^2, \quad a_3 = -a^3. \tag{1.6}$$

We shall adopt the Einstein summation convention, which assumes automatic summation over repeated indices. Thus, (1.5) can be more compactly written as

$$a_\mu = g_{\mu\nu} a^\nu . \tag{1.7}$$

Similarly, indices are raised by

$$a^\mu = g^{\mu\nu} a_\nu\,, \tag{1.8}$$

where $g^{\mu\nu}$ denotes the components of the inverse matrix of $g_{\mu\nu}$. Furthermore, we have

$$g_\mu{}^\nu = g_{\mu\rho} g^{\rho\nu} = g^\mu{}_\nu = \delta_\mu{}^\nu, \tag{1.9}$$

where $\delta_\mu{}^\nu$ is the Kronecker symbol:

$$\delta_\mu{}^\nu = \begin{cases} 1 & \text{if } \mu = \nu; \\ 0 & \text{if } \mu \neq \nu. \end{cases} \tag{1.10}$$

The scalar product of two four-vectors a^μ and b^μ is obtained by contracting the contravariant components of one with the covariant components of the other:

$$a_\mu b^\mu = a^\mu b_\mu = a^0 b^0 - \mathbf{a} \cdot \mathbf{b}\,, \tag{1.11}$$

where $\mathbf{a} \cdot \mathbf{b}$ denotes the usual dot product between three-vectors.

In particular, the norm of a four-vector a^μ is

$$a_\mu a^\mu = (a^0)^2 - \mathbf{a}^2. \tag{1.12}$$

It is *not* positive-definite. Four-vectors can be classed into three types, depending on the sign of their norm:

$$a_\mu a^\mu \begin{cases} < 0 & a^\mu \text{ space-like;} \\ = 0 & a^\mu \text{ null;} \\ > 0 & a^\mu \text{ time-like.} \end{cases} \tag{1.13}$$

This classification corresponds to the position of the vector with respect to the light cone $x_\mu x^\mu = 0$. The two latter cases can be further classified according to the sign of the time component:

$$a^0 \begin{cases} > 0 & a^\mu \text{ points towards the future;} \\ < 0 & a^\mu \text{ points towards the past.} \end{cases} \tag{1.14}$$

We retain the notation

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right). \tag{1.15}$$

The four-partial-differential operator $\frac{\partial}{\partial x^\mu}$ is a *covariant* vector, which we denote by ∂_μ:

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial ct}, \nabla \right). \tag{1.16}$$

The wave operator is defined by

$$\partial^2 = \partial_\mu \partial^\mu = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2. \tag{1.17}$$

This operator is often denoted by $\Box$.

Finally, we introduce the anti-symmetric tensor $\epsilon^{\lambda\mu\nu\rho}$, which is completely anti-symmetric with respect to its four indices. It is equal to 0 if any two indices are equal, $+1$ if $(\lambda\mu\nu\rho)$ is an even permutation of (0123), and -1 if $(\lambda\mu\nu\rho)$ is an odd permutation of (0123).

1.2 Lorentz group

Recall that the Lorentz transformation along the x direction, such that the new (primed) frame moves with velocity $v = \beta c$ with respect to the old one, is

$$x'^0 = \frac{x^0 - \beta x^1}{\sqrt{1-\beta^2}}\,, \qquad x'^1 = \frac{x^1 - \beta x^0}{\sqrt{1-\beta^2}}\,, \tag{1.18}$$

with x_2 and x_3 unchanged. If we set

$$\tanh\omega = \beta\,, \tag{1.19}$$

(1.18) can be rewritten as

$$\begin{aligned} x'^0 &= x^0\cosh\omega - x^1\sinh\omega\,, \\ x'^1 &= -x^0\sinh\omega + x^1\cosh\omega\,. \end{aligned} \tag{1.20}$$

Thus the Lorentz transformation (1.18) can be thought of as a (hyperbolic) rotation in Minkowski space, with angle ω in the 0-1 plane.

More generally, we can write the Lorentz transformation as

$$x'^\mu = a^\mu{}_\nu x^\nu, \tag{1.21}$$

where in the above case,

$$a^\mu{}_\nu = \begin{pmatrix} \cosh\omega & -\sinh\omega & 0 & 0 \\ -\sinh\omega & \cosh\omega & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.22}$$

We can obtain the matrices $a_\mu{}^\nu$, $a^{\mu\nu}$ and $a_{\mu\nu}$ from $a^\mu{}_\nu$ using the metric $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$. The condition that the norm of x^μ is real and invariant under Lorentz transformations implies that

$$\begin{aligned} a_{\mu\nu}{}^* &= a_{\mu\nu}\,, \\ a_{\mu\nu}a^{\mu\lambda} &= a_{\nu\mu}a^{\lambda\mu} = \delta_\nu{}^\lambda. \end{aligned} \tag{1.23}$$

It follows that

$$\det|a^\mu{}_\nu| = \pm 1\,. \tag{1.24}$$

These transformations form a group known as the *complete Lorentz group*. It is the group of real linear transformations conserving the scalar product between four-vectors.

If $a^{00} > 0$, the transformation conserves the sense of time-like vectors, i.e., it conserves the sign of the time component of these vectors. Such a transformation is called orthochronous, and the set of these transformations is called the orthochronous Lorentz group.

If in addition $\det |a^{\mu}{}_{\nu}| = +1$, the transformations also conserves the sense of the Cartesian system in ordinary space, and is known as a proper Lorentz transformation. The set of these transformations is known as the proper Lorentz group $\mathcal{L}_0$, and it contains the identity transformation.

All transformations of the proper Lorentz group can be considered as a succession of infinitesimal transformations starting from the identity. An infinitesimal transformation has the form

$$a_{\mu\nu} = g_{\mu\nu} + \omega_{\mu\nu} \ , \tag{1.25}$$

where $|\omega_{\mu\nu}| \ll 1$. The conditions (1.23) then give

$$\omega_{\mu\nu}{}^* = \omega_{\mu\nu} \ , \qquad \omega_{\mu\nu} = -\omega_{\nu\mu} \ , \tag{1.26}$$

i.e., $\omega_{\mu\nu}$ is a real, anti-symmetric tensor. It is convenient to rewrite (1.25) as

$$a_{\mu\nu} = g_{\mu\nu} + \triangle\omega (I_{\alpha\beta})_{\mu\nu} \ , \tag{1.27}$$

where

$$(I_{\alpha\beta})_{\mu\nu} = g_{\alpha\mu} g_{\beta\nu} - g_{\alpha\nu} g_{\beta\mu} \ , \tag{1.28}$$

are clearly anti-symmetric in μ and ν (and also in α and β). $I_{\alpha\beta}$ is a 4×4 matrix (in space-time: μ and ν label the row and column respectively) of coefficients generating a unit Lorentz rotation in the x^{α}-x^{β} plane. $\triangle\omega$ is an infinitesimal quantity parametrising this rotation. For example, the rotations in the 1-2, 2-3 and 3-1 planes are spatial rotations of angle $\triangle\omega$ about the axes z, x, y respectively. The rotations in the planes 0-1, 0-2 and 0-3 are Lorentz transformations of velocity $\triangle\omega$ in the directions x, y, z respectively.

Let us construct a finite proper transformation from the infinitesimal ones. To be definite, we consider a rotation in the 0-1 plane. The relevant generator is

$$(I_{01})^{\mu}{}_{\nu} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{1.29}$$

which we denote by I for convenience. A finite transformation in this direction is therefore

$$\begin{aligned} x'^{\mu} &= \lim_{N\to\infty}\left(g+\frac{\omega}{N}I\right)^{\mu}{}_{\alpha_1}\left(g+\frac{\omega}{N}I\right)^{\alpha_1}{}_{\alpha_2}\cdots x^{\alpha_N} \\ &= (\exp\omega I)^{\mu}{}_{\nu}x^{\nu} \\ &= (\cosh\omega I+\sinh\omega I)^{\mu}{}_{\nu}x^{\nu} \\ &= (1-I^2+I^2\cosh\omega+I\sinh\omega)^{\mu}{}_{\nu}x^{\nu}, \end{aligned} \tag{1.30}$$

where in the last line, we have used that fact that $I^3 = I$. (1.30) is precisely the proper Lorentz transformation (1.21) with $a^{\mu}{}_{\nu}$ given in (1.22).

In addition to infinitesimal transformations discussed above, we may define discrete transformations known as *reflections*. These are the spatial reflection (also called parity transformation) $x'^{\mu} = Px^{\mu}$:

$$x'^0 = x^0\,, \qquad \mathbf{x}' = -\mathbf{x}\,, \tag{1.31}$$

and the time reflection $x'^{\mu} = Tx^{\mu}$:

$$x'^0 = -x^0\,, \qquad \mathbf{x}' = \mathbf{x}\,. \tag{1.32}$$

The orthochronous Lorentz group is made up of $\mathcal{L}_0$ and $P\mathcal{L}_0$. The complete Lorentz group is formed from $\mathcal{L}_0$, $P\mathcal{L}_0$, $T\mathcal{L}_0$ and $PT\mathcal{L}_0$. It consists of four disconnected sectors whose properties are as follows:

Sector	$\det\lvert a^{\mu}{}_{\nu}\rvert$	a^{00}
$\mathcal{L}_0$	$+1$	> 0
$P\mathcal{L}_0$	-1	> 0
$T\mathcal{L}_0$	-1	< 0
$PT\mathcal{L}_0$	$+1$	< 0

1.3 Klein–Gordon equation

Consider a non-relativistic free particle described by quantum mechanics. Its energy is $E = \frac{\mathbf{p}^2}{2m}$ and it is governed by the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\varphi = \frac{(-i\hbar\nabla)^2}{2m}\,\varphi\,. \tag{1.33}$$

But the relativistic expression for energy is $E = \sqrt{c^2\mathbf{p}^2 + m^2c^4}$. Hence, the relativistic generalisation of the Schrödinger equation appears to be

$$i\hbar\frac{\partial}{\partial t}\varphi = \sqrt{c^2(-i\hbar\nabla)^2 + m^2c^4}\,\varphi\,. \tag{1.34}$$

The right-hand side is cumbersome to deal with because of the square root. Klein and Gordon instead considered the 'square' of this equation:

$$\left(i\hbar\frac{\partial}{\partial t}\right)^2\varphi = c^2\left[(-i\hbar\nabla)^2 + m^2c^2\right]\,\varphi\,. \tag{1.35}$$

In covariant notation, this is

$$(\hbar^2\partial^2 + m^2c^2)\varphi = 0\,. \tag{1.36}$$

In the following, we shall adopt units in which $\hbar = c = 1$, so the Klein–Gordon equation becomes

$$(\partial^2 + m^2)\varphi = 0\,. \tag{1.37}$$

An elementary plane-wave solution of the Klein–Gordon equation is

$$\varphi(x) = A\,\mathrm{e}^{-i\,p\cdot x}, \tag{1.38}$$

with

$$p^2 = m^2. \tag{1.39}$$

This condition is equivalent to

$$E = \pm\sqrt{\mathbf{p}^2 + m^2}\,. \tag{1.40}$$

In order words, there appears to be both positive and negative energy solutions! This is to be compared with the non-relativistic case, where

$$\varphi(\mathbf{x}, t) = A\mathrm{e}^{-i\,Et}\mathrm{e}^{-i\,\mathbf{p}\cdot\mathbf{x}}, \tag{1.41}$$

represents a solution with positive energy. For a free particle with constant energy, this problem can be avoided by choosing the particle to have positive energy, and ignoring the negative energy states. But an interacting particle will exchange energy with its environment, and there is nothing to stop it from cascading down the negative energy states to $E = -\infty$, and emitting an infinite amount of energy in the process.

This is not the only problem of the Klein–Gordon equation. Recall that the probability density for the Schrödinger equation is

$$\rho = \varphi^* \varphi \,, \tag{1.42}$$

and the probability current is

$$\mathbf{j} = -\frac{i\hbar}{2m}(\varphi^* \nabla \varphi - \varphi \nabla \varphi^*) \,. \tag{1.43}$$

They obey the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 \,. \tag{1.44}$$

The corresponding expressions for the Klein–Gordon equation can be determined from the above by demanding relativistic invariance. ρ should not be a scalar, but the time component of a four-vector whose spatial component is $\mathbf{j}$ given in (1.43), i.e.,

$$\rho = \frac{i\hbar}{2m}\left(\varphi^* \frac{\partial \varphi}{\partial t} - \varphi \frac{\partial \varphi^*}{\partial t}\right) . \tag{1.45}$$

The four-current can then be written as

$$j^\mu = (\rho, \mathbf{j}) = \frac{i\hbar}{m}\, \varphi^* (\overleftrightarrow{\partial}_0, -\overleftrightarrow{\nabla})\varphi = \frac{i\hbar}{m}\, \varphi^* \overleftrightarrow{\partial}^\mu \varphi \,, \tag{1.46}$$

where the double-sided derivative is defined by

$$a \overleftrightarrow{\partial}^\mu b \equiv \tfrac{1}{2}[a\, \partial^\mu b - (\partial^\mu a) b] \,. \tag{1.47}$$

It can be checked that the continuity equation, in relativistic notation:

$$\partial_\mu j^\mu = 0 \,, \tag{1.48}$$

is satisfied.

Note that the relativistic probability density (1.45) is not positive definite, unlike the non-relativistic one (1.42) which clearly is. This causes difficulty in interpreting ρ as a probability density. Thus, the interpretation of the Klein–Gordon equation as a single-particle equation, with wave function φ, has to be abandoned.

1.4 Dirac equation

Because of the difficulties inherent in the Klein–Gordon equation, Dirac tried to replace it with a first-order equation:

$$(-i\gamma_\mu \partial^\mu + m)\psi = 0\,, \tag{1.49}$$

where γ_μ are certain algebraic entities to be determined. There are four properties that Dirac demanded of this equation:

1 It must admit plane-wave solutions which satisfy the relativistic energy-mass relation $E^2 = p^2 + m^2$.

2 The corresponding Hamiltonian must be hermitian.

3 There is a four-vector current density whose time component is a positive density.

4 The equation is relativistically covariant.

It turns out that the first and second conditions will determine γ_μ, with the third and fourth conditions following directly. Let us consider these points in turn. Point 4 will be addressed in the following section.

1 To get the right relation between energy and momentum, ψ must satisfy the Klein–Gordon equation. Operate on the left with $(i\gamma_\nu \partial^\nu + m)$ to get

$$(\gamma_\nu \gamma_\mu \partial^\mu \partial^\nu + m^2)\psi = 0\,. \tag{1.50}$$

Because $\partial^\mu \partial^\nu = \partial^\nu \partial^\mu$, this equation is equivalent to

$$\left(\tfrac{1}{2}\{\gamma_\mu, \gamma_\nu\}\partial^\mu \partial^\nu + m^2\right)\psi = 0\,, \tag{1.51}$$

where $\{\cdot, \cdot\}$ denotes the anti-commutator bracket:

$$\{a, b\} \equiv ab + ba\,. \tag{1.52}$$

To get the Klein–Gordon equation, we have to impose the condition

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}\,. \tag{1.53}$$

This, in particular, implies that

$${\gamma_0}^2 = 1, \qquad {\gamma_1}^2 = {\gamma_2}^2 = {\gamma_3}^2 = -1\,. \tag{1.54}$$

2 The Hamiltonian corresponding to the Dirac equation should be hermitian. To derive the Hamiltonian, multiply (1.49) on the left by γ^0 to get

$$i\frac{\partial\psi}{\partial t} = (-i\gamma^0\boldsymbol{\gamma}\cdot\nabla + \gamma^0 m)\psi\,. \tag{1.55}$$

Hence,

$$H = \gamma^0\boldsymbol{\gamma}\cdot\mathbf{p} + \gamma^0 m \equiv \boldsymbol{\alpha}\cdot\mathbf{p} + \beta\, m\,. \tag{1.56}$$

We want $\boldsymbol{\alpha}$, β to be hermitian:

$$\begin{aligned} \gamma^{0\dagger} &= \gamma^0, \\ (\gamma^0\boldsymbol{\gamma})^\dagger &= \gamma^0\boldsymbol{\gamma}\,. \end{aligned} \tag{1.57}$$

The latter condition is equivalent to

$$\boldsymbol{\gamma}^\dagger = \gamma^0\boldsymbol{\gamma}\gamma^0, \qquad \text{or} \qquad \boldsymbol{\gamma}^\dagger = -\boldsymbol{\gamma}\,. \tag{1.58}$$

Together, these relations can be written as

$$\gamma_\mu{}^\dagger = \gamma^0\gamma_\mu\gamma^0. \tag{1.59}$$

Now we can represent the γ^μ explicitly by matrices. The conditions (1.53) and (1.59) imply that they must be at least 4×4 matrices. One possible representation is

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{1.60}$$

where $\boldsymbol{\sigma}$ are the three Pauli matrices. At the same time, ψ is represented as a four-component column vector:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \tag{1.61}$$

The representation (1.60) is known as the Dirac representation. It is not unique. Other representations can be obtained by unitary transformations

$$\gamma'^\mu = U\gamma^\mu U^{-1}, \tag{1.62}$$

where U is a unitary 4×4 matrix. The Dirac equation will then be satisfied by

$$\psi' = U\psi\,. \tag{1.63}$$

3 Let us now construct a probability current j^μ for the Dirac equation. Taking the hermitian conjugate of (1.49) and using (1.59), we have

$$\psi^\dagger(-i\gamma^0\overleftarrow{\partial}_0 + i\boldsymbol{\gamma}\cdot\overleftarrow{\nabla} - m) = 0\,, \tag{1.64}$$

where $\psi^\dagger = (\psi_1{}^*, \psi_2{}^*, \psi_3{}^*, \psi_4{}^*)$ is a four-component row vector. If we define the *adjoint* of ψ to be

$$\bar\psi \equiv \psi^\dagger\gamma^0, \tag{1.65}$$

then (1.64) can be simply written as

$$\bar\psi(i\gamma^\mu\overleftarrow{\partial}_\mu + m) = 0\,. \tag{1.66}$$

Equations (1.49) and (1.64) can now be used to show that the current

$$j^\mu = \bar\psi\gamma^\mu\psi\,, \tag{1.67}$$

is conserved:

$$\begin{aligned}\partial_\mu j^\mu &= (\partial_\mu\bar\psi)\gamma^\mu\psi + \bar\psi\gamma^\mu\partial_\mu\psi \\ &= (im\bar\psi)\psi + \bar\psi(-im\psi) = 0\,.\end{aligned} \tag{1.68}$$

The probability density follows from (1.67):

$$j^0 = \bar\psi\gamma^0\psi = \psi^\dagger\psi\,, \tag{1.69}$$

and is positive definite. Hence, j^0 is fit to serve as the probability density for the particle in question. The Dirac equation has resolved one problem that plagued the Klein–Gordon equation.

1.5 Relativistic covariance

As there is no preferred Lorentz frame, the Dirac equation should look the same in all frames, i.e., be covariant. Under the Lorentz transformation $x'^\mu = a^{\mu\nu}x_\nu$, we have

$$\partial'_\mu = a_{\mu\nu}\partial^\nu. \tag{1.70}$$

Suppose that ψ transforms as

$$\psi'(x') = S\psi(x)\,. \tag{1.71}$$

The Dirac equation in the new frame is then

$$(-i\gamma_\mu\partial'^\mu + m)\psi' = 0\,, \tag{1.72}$$

which implies that

$$S^{-1}\gamma^\mu S = a^{\mu\nu}\gamma_\nu\,. \tag{1.73}$$

This is the fundamental relation defining S. In seeking to prove the relativistic covariance of the Dirac equation, we are seeking a solution to (1.73).

We shall now construct S for an infinitesimal proper Lorentz transformation of the form (1.25). Expanding S in powers of $\omega_{\mu\nu}$ and keeping only the linear term in the infinitesimal generators, we write

$$S = 1 - \tfrac{i}{4}\sigma_{\mu\nu}\omega^{\mu\nu}, \tag{1.74a}$$

and

$$S^{-1} = 1 + \tfrac{i}{4}\sigma_{\mu\nu}\omega^{\mu\nu}. \tag{1.74b}$$

Note that $\sigma_{\mu\nu} = -\sigma_{\nu\mu}$. Each of the six coefficients $\sigma_{\mu\nu}$ is a 4×4 matrix. Inserting (1.25) and (1.74) into (1.73), and keeping first-order terms in $\omega^{\mu\nu}$, we find

$$\omega^\mu{}_\nu\gamma^\nu = -\tfrac{i}{4}\omega^{\rho\sigma}(\gamma^\mu\sigma_{\rho\sigma} - \sigma_{\rho\sigma}\gamma^\mu)\,. \tag{1.75}$$

From the anti-symmetry of $\omega^{\mu\nu}$, it follows that

$$2i(g^\mu{}_\rho\gamma_\sigma - g^\mu{}_\sigma\gamma_\rho) = [\gamma^\mu, \sigma_{\rho\sigma}]\,. \tag{1.76}$$

It can be checked that a solution to this equation is given by

$$\sigma_{\mu\nu} = \tfrac{i}{2}[\gamma_\mu, \gamma_\nu]\,. \tag{1.77}$$

Thus, S for an infinitesimal Lorentz transformation is given by

$$S = 1 + \tfrac{1}{8}[\gamma_\mu, \gamma_\nu]\omega^{\mu\nu}. \tag{1.78}$$

To construct a finite spinor transformation S, we have, as in (1.27),

$$\begin{aligned}\psi'(x') &= \lim_{N\to\infty}\left(1 - \frac{i}{4}\frac{\omega}{N}\sigma_{\mu\nu}(I_{\alpha\beta})^{\mu\nu}\right)^N \psi(x) \\ &= \exp\left(-\tfrac{i}{4}\omega\sigma_{\mu\nu}(I_{\alpha\beta})^{\mu\nu}\right)\psi(x)\,. \end{aligned} \tag{1.79}$$

Specialising again to the transformation (1.29), we get

$$\psi'(x') = \exp\left(-\tfrac{i}{2}\omega\sigma_{01}\right)\psi(x)\,, \tag{1.80}$$

where x' and x are related by (1.22).

Similarly, for a rotation about the z axis,

$$\psi'(x') = \exp\left(-\tfrac{i}{2}\omega\sigma_{12}\right)\psi(x)\,, \tag{1.81}$$

where

$$\sigma_{12} = \begin{pmatrix}\sigma_3 & 0\\ 0 & \sigma_3\end{pmatrix}. \tag{1.82}$$

The appearance of the half-angle in (1.81) is an expression of the double valuedness of the spinor law of rotation: it takes a rotation of 4π radians to return $\psi(x)$ to its original value. Because of this, observables in the Dirac theory must be bilinear, or an even power in $\psi(x)$.

The simplest example of a bilinear is $\bar{\psi}\psi$. To find out how it transforms under a Lorentz transformation, we first have to find out how $\bar{\psi}$ transforms. Taking the hermitian conjugate of (1.73), and using (1.59) and the fact that $a^{\mu\nu}$ are real,

$$S^\dagger\gamma^0\gamma^\mu\gamma^0 S^{-1\dagger} = a^{\mu\nu}{\gamma_\nu}^\dagger. \tag{1.83}$$

Hence

$$\gamma^0 S^\dagger\gamma^0\gamma^\mu\gamma^0 S^{-1\dagger}\gamma^0 = a^{\mu\nu}\gamma_\nu = S^{-1}\gamma^\mu S\,, \tag{1.84}$$

or

$$(S\gamma^0 S^\dagger\gamma^0)\gamma^\mu(S\gamma^0 S^\dagger\gamma^0)^{-1} = \gamma^\mu. \tag{1.85}$$

Thus, we have the commutator

$$[S\gamma^0 S^\dagger\gamma^0, \gamma^\mu] = 0\,, \tag{1.86}$$

i.e., $S\gamma^0 S^\dagger \gamma^0$ commutes with all the γ-matrices and therefore with all 4×4 matrices. By Schur's lemma, it must be a constant multiple of the unit matrix:

$$S\gamma^0 S^\dagger \gamma^0 = b\,, \tag{1.87}$$

which is real. If we choose to normalise $S\psi$ by requiring that $\det S = 1$, then we have $b^4 = 1$ or $b = \pm 1$. We pick the positive sign because we are only interested in proper Lorentz transformations.

With this, observe that

$$\bar{\psi}' = \psi'^\dagger \gamma^0 = (S\psi)^\dagger \gamma^0 = \bar{\psi}\gamma^0 S^\dagger \gamma^0 = \bar{\psi} S^{-1}. \tag{1.88}$$

Thus, under a Lorentz transformation,

$$\bar{\psi}'\psi' = \bar{\psi} S^{-1} S \psi = \bar{\psi}\psi\,, \tag{1.89}$$

i.e., $\bar{\psi}\psi$ is a scalar quantity.

To check that the probability current for the Dirac equation transforms like a vector under a Lorentz transformation, we observe that

$$\bar{\psi}'\gamma^\mu\psi' = \bar{\psi} S^{-1}\gamma^\mu S\psi = a^{\mu\nu}(\bar{\psi}\gamma_\nu\psi)\,, \tag{1.90}$$

i.e., $\bar{\psi}\gamma^\mu\psi$ is a four-vector. Similarly, it can be checked that $\bar{\psi}\sigma^{\mu\nu}\psi$ transforms as a second-rank, anti-symmetric tensor.

At this point, we introduce a hermitian 4×4 matrix

$$\gamma_5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3\,, \tag{1.91}$$

which has the remarkable property that it anti-commutes with every one of the γ^μ, $\mu = 0, 1, 2, 3$:

$$\{\gamma^\mu, \gamma_5\} = 0\,. \tag{1.92}$$

Also note that

$${\gamma_5}^2 = 1\,. \tag{1.93}$$

Explicitly, we have

$$\gamma_5 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \tag{1.94}$$

in the Dirac representation. It can be checked that under a proper Lorentz transformation S,

$$S^{-1}\gamma_5 S = \gamma_5 \,, \tag{1.95}$$

but under a parity transformation P

$$P^{-1}\gamma_5 P = -\gamma_5 \,. \tag{1.96}$$

Hence, $\bar{\psi}\gamma_5\psi$ transforms exactly like the scalar $\bar{\psi}\psi$ under proper Lorentz transformations, but changes its sign under space inversion. This type of transformation is characteristic of a *pseudoscalar*. Similarly, $\bar{\psi}\gamma_5\gamma^\mu\psi$ transforms in the same way as $\bar{\psi}\gamma^\mu\psi$ under proper Lorentz transformations, but acquires a minus sign under space inversion. This is expected of an *axial vector* or pseudovector.

The question naturally arises: have we listed all possible bilinear covariants of the form $\bar{\psi}\Gamma\psi$? It can be shown that the matrices

$$\Gamma^S = 1\,, \quad \Gamma^V_\mu = \gamma_\mu\,, \quad \Gamma^T_{\mu\nu} = \sigma_{\mu\nu}\,, \quad \Gamma^P = \gamma_5\,, \quad \Gamma^A_\mu = \gamma_5\gamma_\mu\,, \tag{1.97}$$

are linearly independent and form a basis of all 4×4 matrices. Hence, we have found all the possible bilinear covariants.

1.6 Free particle solutions and their interpretation

Let us look for simple plane-wave solutions of the Dirac equation. It is convenient to work in momentum space by assuming a solution of the form

$$\psi(x) = u(p)\mathrm{e}^{-i\,p\cdot x}. \tag{1.98}$$

The Dirac equation then becomes

$$(\gamma\cdot p - m)u(p) = 0\,. \tag{1.99}$$

We multiply on the left by $(\gamma\cdot p + m)$ and use the relation $(\gamma\cdot p)^2 = p^2$ to get

$$(p^2 - m^2)u(p) = 0\,. \tag{1.100}$$

Thus $u(p)$ is non-zero only if

$$E = \pm\sqrt{\mathbf{p}^2 + m^2}\,. \tag{1.101}$$

Like the Klein–Gordon equation, the Dirac equation admits solutions with negative energy. We will see below how Dirac interpreted these solutions as anti-particles.

Now for a particle at rest with $\mathbf{p} = 0$, we have

$$(\pm m\gamma^0 - m)u(p) = 0\,. \tag{1.102}$$

In the Dirac representation (1.60), there are two solutions to this equation

$$u_1 = N_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad u_2 = N_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \tag{1.103}$$

with positive energy $E = +m$, and two solutions

$$v_1 = N_1' \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad v_2 = N_2' \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{1.104}$$

with negative energy $E = -m$. Here, N_1, N_2, N_1' and N_2' are normalisation constants.

Away from $\mathbf{p} = 0$, it can be checked that

$$u(p) = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \chi \\ \dfrac{\boldsymbol{\sigma}\cdot\mathbf{p}\,\chi}{E+m} \end{pmatrix}, \qquad v(p) = \sqrt{\frac{-E+m}{2m}} \begin{pmatrix} \dfrac{-\boldsymbol{\sigma}\cdot\mathbf{p}\,\chi}{-E+m} \\ \chi \end{pmatrix}, \tag{1.105}$$

where

$$\chi = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.106}$$

and the normalisation has been chosen so that $\bar{u}u = 1 = -\bar{v}v$.

These solutions to the Dirac equation are not only specified by the continuous momentum (or coordinate) label, but also by the discontinuous label r, $r = 1, 2$. These new degrees of freedom are related to the intrinsic spin of the particle, as the following argument shows.

In contrast to the non-relativistic Schrödinger equation, the Dirac Hamiltonian operator does not commute with the components of the orbital angular momentum operator $\mathbf{L} = \mathbf{x} \times \mathbf{p} = \mathbf{x} \times (-i\nabla)$. Using (1.56), we have, for example, that

$$[H, L^1] = \alpha^3 \frac{\partial}{\partial x^2} - \alpha^2 \frac{\partial}{\partial x^3} \neq 0\,. \tag{1.107}$$

So $\mathbf{L}$ is not a constant of motion. To save the law of angular momentum conservation, we must amend the orbital momentum by an additional intrinsic spin angular momentum $\mathbf{S}$, given by

$$\mathbf{S} = \frac{1}{2}\begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}. \tag{1.108}$$

The eigenvalues of the third spin component are $\pm\frac{1}{2}$. Thus, the Dirac equation describes particles of spin $\frac{1}{2}$, such as the electron. The spin has been automatically incorporated into the wave equation through the increase in the number of components of ψ, and for this reason, ψ is known as a 'spinor.'

The solutions u_1 and u_2 in (1.103) therefore describes particles with spin $\frac{1}{2}$ and $-\frac{1}{2}$ respectively, being eigenstates of S_3. In a general moving frame, the plane-wave solution are eigenstates of $\mathbf{S}\cdot\mathbf{p}/|\mathbf{p}|$. This operator commutes with the Hamiltonian and is known as the helicity operator. The helicity eigenstates with eigenvalues $+1$ and -1 are referred to, respectively, as the right-handed state (spin parallel to motion) and the left-handed state (spin opposite to motion).

We still have to deal with the negative energy states. Dirac attempted to remedy this difficulty by proposing the so-called hole theory. Although it is no longer accepted as correct, it nevertheless led to the prediction of anti-particles, and we shall review it here.

Because spin-$\frac{1}{2}$ particles, electrons say, obey the Pauli exclusion principle, Dirac suggested that the ground state already has all the possible negative energy levels occupied, so that normal electrons cannot descend into the negative energy states. If we supply an energy $E \geq 2m$, for example in the form of a gamma ray, we can knock an electron out of one of the negative energy states into one of the positive ones. At the same time, a 'hole' will be produced in the 'sea' of negative energy electrons. It can be shown that the absence of a particle with *negative energy, negative charge and momentum* $\mathbf{p}$, will look like the presence of an anti-particle with *positive energy, positive charge and momentum* $-\mathbf{p}$. Thus, the overall effect of the (radiative) energy is to create an electron-positron pair.

$u_r(p)$, $r = 1, 2$ therefore represent the two spin states of a particle (electron), while $v_r(p)$ the two spin states of the corresponding anti-particle (positron). It is convenient to redefine

$$v_r(p) \to v_r(-p)\,, \tag{1.109}$$

so that now p is the 4-momentum of the antiparticle with $p^0 > 0$.

Define

$$\Lambda_\pm = \frac{m \pm \gamma \cdot p}{2m} \,. \tag{1.110}$$

They are projection operators as they satisfy the relations

$$\Lambda_\pm^2 = \Lambda_\pm \,, \quad \Lambda_+ + \Lambda_- = 1 \,, \quad \Lambda_+\Lambda_- = 0 = \Lambda_-\Lambda_+ \,. \tag{1.111}$$

In fact, they project onto positive and negative energy states respectively:

$$\begin{aligned} \Lambda_+(p)v_r(p) &= 0 \,, \\ \Lambda_-(p)u_r(p) &= 0 \,. \end{aligned} \tag{1.112}$$

These positive and negative energy states are mutually orthogonal:

$$\begin{aligned} \bar{u}_r(p)u_s(p) = 2m\delta_{rs} &= -\bar{v}_r(p)v_s(p) \,, \\ \bar{v}_r(p)u_s(p) &= 0 \,, \end{aligned} \tag{1.113}$$

and form a complete basis of solutions:

$$\begin{aligned} \sum_{r=1,2} u_r(p)\bar{u}_r(p) &= 2m\Lambda_+(p) \,, \\ \sum_{r=1,2} v_r(p)\bar{v}_r(p) &= -2m\Lambda_-(p) \,, \end{aligned} \tag{1.114}$$

and

$$\sum_{r=1,2} (u_r\bar{u}_r - v_r\bar{v}_r) = 2m \,. \tag{1.115}$$

1.7 Interaction with electromagnetic field

We now wish to briefly discuss how the Dirac equation can be extended to the case where the relativistic particle is not free, but moves in an external electromagnetic field. This, in particular, would describe an electron in the Coulomb field of the atomic nucleus.

Recall from classical electromagnetism that the electromagnetic field is specified by the (three-) vector potential $\mathbf{A}(\mathbf{x}, t)$ and the scalar potential $\phi(\mathbf{x}, t)$. Its effect is to modify the momentum of a particle with electric charge q by

$$\mathbf{p} \to \mathbf{p} - q\mathbf{A} \,, \tag{1.116}$$

and the energy by

$$E \to E - q\phi \,. \tag{1.117}$$

In relativistic notation, we can combine the electromagnetic potentials into a four-vector potential $A^\mu = (\phi, \mathbf{A})$. The modifications (1.116) and (1.117) can then be written as

$$p^\mu \to p^\mu - qA^\mu(p) \,, \tag{1.118}$$

or as a quantum mechanical differential operator

$$i\partial^\mu \to i\partial^\mu - qA^\mu(x) \,. \tag{1.119}$$

The Dirac equation (1.49) is therefore modified to

$$[-\gamma \cdot (i\partial - qA) + m]\psi = 0 \,, \tag{1.120}$$

in the presence of an electromagnetic field.

Let us, as an example, apply (1.120) to calculate the energy of an electron in an external magnetic field B. We premultiply (1.120) by $\gamma \cdot (i\partial - qA) + m$, and after some algebra, we obtain

$$\left\{ \left(-(i\partial - qA)^2 + m^2 \right) + \tfrac{1}{4} iq[\gamma_\mu, \gamma_\nu]F^{\mu\nu} \right\} \psi = 0 \,. \tag{1.121}$$

Here,

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu, \tag{1.122}$$

is the electromagnetic field strength tensor, such that

$$\begin{aligned} F^{0i} &= E^i, \\ F^{ij} &= \epsilon^{ijk} B^k, \qquad i, j, k = 1, 2, 3 \,. \end{aligned} \tag{1.123}$$

The first term in (1.121) is what would have been obtained from the Klein–Gordon equation in an external electromagnetic field. The second term is peculiar to the Dirac equation, i.e. to particles of spin $\frac{1}{2}$.

We shall go over to the low-energy, non-relativistic limit. In this case,

$$E \simeq m + E_{\mathrm{NR}}, \tag{1.124}$$

with $E_{\rm NR} \equiv \frac{\mathbf{p}^2}{2m} \ll m$. We also assume that $|q\phi| \ll m$. Then

$$(p - qA)^2 - m^2 \simeq 2m(E_{\rm NR} - q\phi) - (\mathbf{p} - q\mathbf{A})^2. \tag{1.125}$$

Now, the tensor $\sigma^{\mu\nu} \equiv \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ has elements (in the Dirac representation)

$$\sigma^{0i} = i \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix}, \qquad \sigma^{ij} = \epsilon^{ijk}\sigma^k. \tag{1.126}$$

Since there is no electric field assumed to be present, σ^{0i} does not contribute in (1.121). Combining (1.121) and (1.125), we finally have

$$E_{\rm NR}\psi = \left\{ \frac{1}{2m}(\mathbf{p} - q\mathbf{A})^2 + q\phi - \frac{q}{2m}\boldsymbol{\sigma} \cdot \mathbf{B} \right\} \psi\,. \tag{1.127}$$

The last term corresponds to the interaction of the $\mathbf{B}$ field with a magnetic dipole of moment

$$-\frac{q\boldsymbol{\sigma}}{2m} = -\frac{q}{2m} g\mathbf{S}\,, \tag{1.128}$$

with $g = 2$. Higher-order corrections give

$$g = 2 \times 1.001596522\,, \tag{1.129}$$

which is in agreement with experiment to one part in 10^{10}. This is a spectacular confirmation of QED.

One can also solve the hydrogen atom using the Dirac equation. Because $\mathbf{L}$ is no longer a good quantum number, the degeneracy of the non-relativistic levels is resolved. This leads to the prediction of hyperfine structure and the Lamb shift, by QED.

1.8 References and further reading

Relativistic quantum mechanics, and in particular the Dirac equation, is treated in most books on advanced quantum mechanics and quantum field theory. Good references include Bjorken and Drell, *Relativistic Quantum Mechanics*; Messiah, *Quantum Mechanics, Vol. II*; and Sakurai, *Advanced Quantum Mechanics*.