The Dirac Equation and the Lorentz Group

Joel G. Broida
University of Colorado, Boulder

February 16, 2009

Contents

The Dirac Equation and The Lorentz Group

Part I – Classical Approach

1 Derivation of the Dirac Equation

The basic idea is to use the standard quantum mechanical substitutions

$$\mathbf{p} \to -i\hbar\boldsymbol{\nabla} \qquad \text{and} \qquad E \to i\hbar\frac{\partial}{\partial t} \tag{1}$$

to write a wave equation that is first-order in both E and $\mathbf{p}$. This will give us an equation that is both relativistically covariant and conserves a positive definite probability density.

We start by assuming that we can factor the relativistic expression $E^2 = p^2 + m^2$ into the form (we will use units with $\hbar = c = 1$ from now on)

$$E = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m \tag{2}$$

where $\boldsymbol{\alpha}$ and β are to be determined. Note that $\boldsymbol{\alpha}$ and β can not simply be numbers because equation (2) would not even be rotationally invariant. Since we must still satisfy $E^2 = p^2 + m^2$, we have

$$\begin{aligned}
E^2 &= (\boldsymbol{\alpha} \cdot \mathbf{p} + \beta m)^2 = (\alpha^i p_i + \beta m)(\alpha^j p_j + \beta m) \\
&= \alpha^i \alpha^j p_i p_j + (\alpha^i \beta + \beta \alpha^i) p_i m + \beta^2 m^2 \\
&= \frac{1}{2}(\alpha^i \alpha^j + \alpha^j \alpha^i) p_i p_j + (\alpha^i \beta + \beta \alpha^i) p_i m + \beta^2 m^2
\end{aligned}$$

where we used the fact that $p_i p_j = p_j p_i$. This requires that

$$\frac{1}{2}(\alpha^i \alpha^j + \alpha^j \alpha^i) = \delta^{ij} \tag{3a}$$

$$\alpha^i \beta + \beta \alpha^i = 0 \tag{3b}$$

$$\beta^2 = I \tag{3c}$$

Since pure numbers commute, let us assume that the α^i and β are matrices. Using equations (3), we define the matrices

$$\gamma^i := \beta \alpha^i \qquad \text{and} \qquad \gamma^0 := \beta.$$

Then

$$\begin{aligned}
2\delta^{ij} &= \alpha^i \alpha^j + \alpha^j \alpha^i = \beta^2 \alpha^i \alpha^j + \beta^2 \alpha^j \alpha^i = -\beta \alpha^i \beta \alpha^j - \beta \alpha^j \beta \alpha^i \\
&= -(\gamma^i \gamma^j + \gamma^j \gamma^i)
\end{aligned}$$

and

$$0 = \alpha^i \beta + \beta \alpha^i \implies 0 = \beta \alpha^i \beta + \beta^2 \alpha^i = \gamma^i \gamma^0 + \gamma^0 \gamma^i$$

and hence we have

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = \{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \tag{4}$$

where we are using the metric $g = \operatorname{diag}(1, -1, -1, -1)$, i.e.,

$$(g_{\mu\nu}) = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix} = (g^{\mu\nu}).$$

Matrices satisfying equation (4) are said to form a **Clifford algebra**. Note in particular that we also have

$$(\alpha^i)^2 = I.$$

From equation (3b) we see that $\beta \alpha^i \beta = -\alpha^i$ and $\alpha^i \beta \alpha^i = -\beta$. Using the cyclic property of the trace along with $\beta^2 = (\alpha^i)^2 = I$, these imply that $\operatorname{tr} \beta = \operatorname{tr} \alpha^i = 0$. Now let λ be an eigenvalue of β. Then $\beta v = \lambda v$ implies $v = \beta^2 v = \lambda \beta v = \lambda^2 v$ and therefore $\lambda = \pm 1$. But the trace of a matrix is the sum of its eigenvalues (i.e., if $P^{-1}AP = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then $\operatorname{tr} A = \operatorname{tr} P^{-1}AP = \operatorname{tr} D$), and hence it follows that β must be even-dimensional, with an exactly analogous result for α^i.

Now, the energy operator E must have real eigenvalues and hence must be Hermitian. Since $\mathbf{p}$ is already Hermitian, it follows from equation (2) that $\boldsymbol{\alpha}$ and β must be Hermitan matrices. The most general 2×2 Hermitian matrix is of the form

$$\begin{bmatrix} z & x - iy \\ x + iy & t \end{bmatrix} = x\sigma_1 + y\sigma_2 + \frac{z}{2}(\sigma_3 + I) - \frac{t}{2}(\sigma_3 - I)$$

where

$$\sigma_1 = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \qquad \sigma_2 = \begin{bmatrix} & -i \\ i & \end{bmatrix} \qquad \sigma_3 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}.$$

Hence the most general 2×2 Hermitian matrix is a linear combination of the three (Hermitan) Pauli matrices and the identity matrix. (It is easy to see that if we write $\sigma_0 := I$, then

$$0 = \sum_{i=0}^{3} c_i \sigma_i = \begin{bmatrix} c_0 + c_3 & c_1 - ic_2 \\ c_1 + ic_2 & c_0 - c_3 \end{bmatrix}$$

implies that all of the c_i must equal zero, and hence the four matrices $\sigma_0, \ldots, \sigma_3$ are linearly independent and form a basis for the space $M_2(\mathbb{C})$.) If we take the α's to be linear combinations of the σ's, then this leaves $\beta = I$. But I commutes with everything, so it certainly can't anticommute with the α's. Thus we assume that the γ's are in fact 4×4 matrices.

In block matrix form, we define the **standard representation** to be

$$\beta = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}. \tag{5a}$$

Since the α's are Hermitan, we have

$$\begin{aligned} 0 = \alpha^i \beta + \beta \alpha^i &= \begin{bmatrix} A & B \\ B^\dagger & C \end{bmatrix} \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} + \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \begin{bmatrix} A & B \\ B^\dagger & C \end{bmatrix} \\ &= \begin{bmatrix} 2A & \\ & -2C \end{bmatrix} \end{aligned}$$

so that $A = C = 0$ and we can choose

$$\boldsymbol{\alpha} = \begin{bmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{bmatrix}. \tag{5b}$$

In other words, we take the standard representation of the gamma matrices to be (in block matrix form)

$$\gamma^0 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad\qquad \boldsymbol{\gamma} = \begin{bmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{bmatrix}. \tag{6}$$

I leave it as an exercise to show directly that if $\{\gamma^\mu\}$ is a set of matrices satisfying $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$, then $\gamma^\mu \neq \gamma^\nu$ for $\mu \neq \nu$ and the γ's are linearly independent.

Next, recall that the gradient is defined by $\boldsymbol{\nabla} = \partial/\partial \mathbf{x}$ so $\nabla^i = \partial/\partial x^i = \partial_i$ and we write

$$\partial^\mu = (\partial^0, \partial^i) = (\partial^0, -\partial_i) = (\partial^0, -\boldsymbol{\nabla})$$

along with $\partial_\mu = (\partial_0, +\boldsymbol{\nabla})$. The quantum mechanical operators are $E = i\partial_0$ and $p^i = -i\partial_i$ or $\mathbf{p} = -i\boldsymbol{\nabla}$, and hence we can write

$$p^\mu = +i\partial^\mu. \tag{7}$$

Then the operators in the Dirac equation become $i\partial^0 = -i\alpha^i\partial_i + \beta m$ so that multiplying through by γ^0 this is $i\gamma^0\partial_0 = -\gamma^i\partial_i + m$ (the I multiplying the m is understood) or simply $i\gamma^\mu\partial_\mu - m = 0$. As a very convenient notational device, we introduce the "Feynman slash" notation for the contraction of any 4-vector a^μ with the gamma matrices γ^μ in which we write

$$\not{a} := \gamma^\mu a_\mu.$$

Using this notation, the **Dirac equation** is then written as

$$\left(i\gamma^\mu \frac{\partial}{\partial x^\mu} - m \right) \psi(x) = 0$$

or simply

$$(i\not{\partial} - m)\psi(x) = 0. \tag{8}$$

Equivalently, we can write this in the form

$$(\not{p} - m)\psi(x) = 0. \tag{9}$$

It is extremely important to realize that now the wavefunction ψ is a 4-component column vector, generally referred to as a **Dirac spinor**. We will see that these four degrees of freedom allow us to describe both positive and negative energy solutions, each with spin 1/2 either up or down. The negative energy solutions are interpreted as describing positive energy antiparticles. In other words, the Dirac equation describes spin 1/2 electrons and positrons (as well as the other leptons and quarks).

Note that $\gamma^{0\dagger} = \gamma^0$ and $\gamma^{i\dagger} = (\beta\alpha^i)^\dagger = \alpha^{i\dagger}\beta^\dagger = \alpha^i\beta = \beta^2\alpha^i\beta = \beta\gamma^i\beta = \gamma^0\gamma^i\gamma^0$ and hence in general we have the very useful result

$$\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0 \tag{10}$$

which is independent of the representation of the gamma matrices. We will see that rather than $\psi^\dagger$, it turns out that the useful quantity will be

$$\overline{\psi} := \psi^\dagger\gamma^0.$$

Taking the adjoint of (8) yields

$$0 = \psi^\dagger(-i\gamma^{\mu\dagger}\overleftarrow{\partial}_\mu - m) = \psi^\dagger(-i\gamma^0\gamma^\mu\gamma^0\overleftarrow{\partial}_\mu - \gamma^0\gamma^0 m)$$

where the symbol $\overleftarrow{\partial}_\mu$ means that the derivative acts to the left. Multiplying this from the right by γ^0 we have $\psi^\dagger\gamma^0(-i\gamma^\mu\overleftarrow{\partial}_\mu - m) = 0$ and hence $\overline{\psi}$ satisfies the equation

$$\overline{\psi}(x)(i\overleftarrow{\not{\partial}} + m) = 0. \tag{11}$$

To get a probability current, we multiply (8) from the left by $\overline{\psi}$ and (11) from the right by ψ and add to obtain

$$i(\overline{\psi}\gamma^\mu\partial_\mu\psi + \partial_\mu\overline{\psi}\gamma^\mu\psi) = 0$$

or simply

$$\partial_\mu(\overline{\psi}\gamma^\mu\psi) = 0. \tag{12}$$

Hence the probability current is given by

$$j^\mu = \overline{\psi}\gamma^\mu\psi$$

and it satisfies the continuity equation $\partial_\mu j^\mu = 0$.

2 Basic Properties of the Dirac Equation

Before we turn to the issue of covariance under Lorentz transformations, let us take a look at some of the basic properties of the Dirac equation.

To begin with, note that equation (8) has solutions of the form

$$\psi(\mathbf{x},t) = u(p)e^{-ip^\mu x_\mu}$$

where $u(p)$ is a 4-component spinor that must satisfy

$$(\not{p} - m)u(p) = 0.$$

This is a set of four homogeneous linear equations, and it will have a nontrivial solution if and only if the matrix $(\not{p} - m)$ does not have an inverse. From equation (4) we see that in general for any 4-vectors a_μ, b_μ we have

$$\not{a}\not{b} + \not{b}\not{a} = (\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu)a_\mu b_\nu = 2a \cdot b$$

so that $\not{p}\not{p} = p^2$. It is then easy to see

$$(\not{p} + m)(\not{p} - m) = p^2 - m^2$$

so that a formal inverse to $(\not{p} - m)$ is $(p^2 - m^2)^{-1}(\not{p} + m)$. But if this inverse is *not* to exist, we must have $p^2 - m^2 = 0$ so that $(p^0)^2 - \mathbf{p}^2 = m^2$ or

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

In other words, the Dirac equation allows solutions with negative energy, and free particles have an energy E with $|E| \geq m$.

Since negative energy states have never been observed, we have to somehow explain their absence. (Such states would have an acceleration in a direction opposite to the applied force. If a particle is accelerated from rest to an energy $E = \int \mathbf{F} \cdot d\mathbf{r} = \pm m \int \mathbf{a} \cdot \mathbf{v}\, dt = \pm m \int (d\mathbf{v}/dt) \cdot \mathbf{v}\, dt = \pm(m/2) \int (dv^2/dt)\, dt < 0$, then we must have $\mathbf{F} = -m\mathbf{a}$.) While the completely correct answer lies in the formalism of relativistic quantum field theory, at the time Dirac postulated that all negative energy states were already filled by an infinite sea of negative energy electrons, and the Pauli principle prevented any positive energy electron from falling down into the negative sea. If such a negative energy electron were hit by a sufficiently energetic photon, it could make the transition to a positive energy state, leaving behind a "hole" that we would perceive as a positive energy positively charged electron, a "positron."

In any case, what can we say about the constants of the motion? Defining the Dirac Hamiltonian

$$H_D = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m$$

we can write the Dirac equation as

$$H_D\psi = i\frac{\partial\psi}{\partial t}$$

which is of the same form as the Schrödinger equation. Then this has a formal solution with time dependence that goes as $e^{-iH_D t}$, and we can define operators

$\mathcal{O}$ in the Heisenberg picture with the usual equation of motion that allows us to look for conserved quantities:

$$\frac{d\mathcal{O}}{dt} = -i[\mathcal{O}, H_D].$$

Let us first look at the orbital angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. Using the commutator identity

$$[ab, c] = a[b, c] + [a, c]b$$

along with the fundamental commutation relations $[p_i, p_j] = 0$ and $[x_i, p_j] = i\delta_{ij}$, we compute (using a sloppy summation convention)

$$[L_i, H_D] = \varepsilon_{ijk}[x^j p^k, H_D] = \varepsilon_{ijk}(x^j[p^k, H_D] + [x^j, H_D]p^k).$$

But

$$[p^k, H_D] = [p^k, \alpha^l p_l + \beta m] = 0$$

while

$$\varepsilon_{ijk}[x^j, H_D]p^k = \varepsilon_{ijk}[x^j, \alpha^l p_l + \beta m]p^k = i\varepsilon_{ijk}\delta^j_l \alpha^l p^k = i(\boldsymbol{\alpha} \times \mathbf{p})_i$$

so that

$$[\mathbf{L}, H_D] = i(\boldsymbol{\alpha} \times \mathbf{p}).$$

This shows that the orbital angular momentum is *not* a constant of the motion.

Now consider the matrix operator

$$\boldsymbol{\sigma}' = \begin{bmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{bmatrix}$$

where the Pauli matrices obey the relations

$$[\sigma_i, \sigma_j] = 2i\varepsilon_{ijk}\sigma_k \tag{13a}$$

$$\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k \tag{13b}$$

and therefore also $(\sigma_i)^2 = 1$. These show that the operator $\mathbf{S}$ defined by $\mathbf{S} = \boldsymbol{\sigma}'/2$ satisfies

$$[S_i, S_j] = i\varepsilon_{ijk}S_k$$

and hence is an angular momentum operator. Since $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma} \cdot \boldsymbol{\sigma} = 3$, we see that

$$\mathbf{S}^2 = \frac{3}{4}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so that $s(s+1) = 3/4$ implies that $s = 1/2$. Thus $\mathbf{S}$ is the spin operator for a particle of spin 1/2. However, we still haven't connected this to the Dirac equation.

Recall that the standard representation for $\boldsymbol{\alpha}$ and β is

$$\boldsymbol{\alpha} = \begin{bmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{bmatrix} \qquad \text{and} \qquad \beta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

so it it easy to see that

$$[\sigma'_i, \alpha_j] = 2i\varepsilon_{ijk}\alpha_k \qquad \text{and} \qquad [\sigma'_i, \beta] = 0.$$

Hence we find that

$$[\sigma'_i, H_D] = [\sigma'_i, \alpha_j p^j + \beta m] = 2i\varepsilon_{ijk}\alpha_k p^j = 2i(\mathbf{p} \times \boldsymbol{\alpha})_i$$

or, alternatively,

$$[\mathbf{S}, H_D] = -i(\boldsymbol{\alpha} \times \mathbf{p}).$$

Combining this with our previous result for $\mathbf{L}$ we see that the operator

$$\mathbf{J} := \mathbf{L} + \mathbf{S}$$

is conserved because $[\mathbf{J}, H_D] = 0$, and furthermore it is an angular momentum operator because

$$[J_i, J_j] = i\varepsilon_{ijk}J_k.$$

I leave it as an exercise to show that $[\mathbf{J}^2, H_D] = [\mathbf{S}^2, H_D] = 0$, and hence the operators $H_D, \mathbf{J}, \mathbf{J}^2$ and $\mathbf{S}^2$ are a mutually commuting set. This then shows that the Dirac equation represents a particle with conserved total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and with spin equal to 1/2.

Now let's take a look at the interaction of a Dirac particle with the electromagnetic field. Since quantum mechanics is formulated using the Hamiltonian, we need to know what the canonical momentum is for a particle of charge e in an electromagnetic field. By definition, this is $p = \partial L/\partial \dot{q}$ where $L = L(q, \dot{q}, t)$ is the Lagrangian of the system. In this case, the answer is we make the replacements

$$\mathbf{p} \to \mathbf{p} - e\mathbf{A} \qquad \text{and} \qquad E \to E - e\phi \tag{14}$$

where $\mathbf{A}$ is the magnetic vector potential and ϕ is the electric potential. For those who are interested, let me somewhat briefly go through the derivation of this result.

In a proper derivation of the Lagrange equations of motion, one starts from d'Alembert's principle and derives Lagrange's equation

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = Q_i \tag{15}$$

where $T = T(q_i, \dot{q}_i)$ is the kinetic energy and Q_i is a generalized force. In the particular case that Q_i is derivable from a conservative force, then we have $Q_i = -\partial V/\partial q_i$. Since the potential energy V is assumed to be independent of $\dot{q}_i$, we can replace $\partial T/\partial \dot{q}_i$ by $\partial (T - V)/\partial \dot{q}_i$ and we arrive at the usual Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \tag{16}$$

where $L = T - V$. However, even if there is no potential function V, we can still arrive at this result if there exists a function $U = U(q_i, \dot{q}_i)$ such that the generalized forces may be written as

$$Q_i = -\frac{\partial U}{\partial \dot{q}_i} + \frac{d}{dt}\frac{\partial U}{\partial \dot{q}_i}$$

because defining $L = T - U$ we again arrive at equation (16). The function U is called a **generalized potential** or a **velocity dependent potential**. We now seek such a function to describe the force on a charged particle in an electromagnetic field.

Recall from electromagnetism that the Lorentz force law is given by

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

or

$$\mathbf{F} = e\Big(-\boldsymbol{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t} + \mathbf{v} \times (\boldsymbol{\nabla} \times \mathbf{A})\Big)$$

where $\mathbf{E} = -\boldsymbol{\nabla}\phi - \partial\mathbf{A}/\partial t$ and $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. Our goal is to write this in the form

$$F_i = -\frac{\partial U}{\partial x_i} + \frac{d}{dt}\frac{\partial U}{\partial \dot{x}_i}$$

for a suitable U. All it takes is some vector algebra. We have

$$\begin{aligned}[\mathbf{v} \times (\boldsymbol{\nabla} \times \mathbf{A})]_i &= \varepsilon_{ijk}\varepsilon^{klm} v^j \partial_l A_m = (\delta_i^l \delta_j^m - \delta_i^m \delta_j^l) v^j \partial_l A_m \\ &= v^j \partial_i A_j - v^j \partial_j A_i = v^j \partial_i A_j - (\mathbf{v} \cdot \boldsymbol{\nabla}) A_i.\end{aligned}$$

But x^i and $\dot{x}^j$ are independent variables (in other words, $\dot{x}^j$ has no explicit dependence on x^i) so that

$$v^j \partial_i A_j = \dot{x}^j \frac{\partial A_j}{\partial x^i} = \frac{\partial}{\partial x^i}(\dot{x}^j A_j) = \frac{\partial}{\partial x^i}(\mathbf{v} \cdot \mathbf{A})$$

and we have

$$[\mathbf{v} \times (\boldsymbol{\nabla} \times \mathbf{A})]_i = \frac{\partial}{\partial x^i}(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \boldsymbol{\nabla}) A_i.$$

But we also have

$$\frac{dA_i}{dt} = \frac{\partial A_i}{\partial x^j}\frac{dx^j}{dt} + \frac{\partial A_i}{\partial t} = v^j \frac{\partial A_i}{\partial x^j} + \frac{\partial A_i}{\partial t} = (\mathbf{v} \cdot \boldsymbol{\nabla}) A_i + \frac{\partial A_i}{\partial t}$$

so that

$$(\mathbf{v} \cdot \boldsymbol{\nabla}) A_i = \frac{dA_i}{dt} - \frac{\partial A_i}{\partial t}$$

and therefore

$$[\mathbf{v} \times (\boldsymbol{\nabla} \times \mathbf{A})]_i = \frac{\partial}{\partial x^i}(\mathbf{v} \cdot \mathbf{A}) - \frac{dA_i}{dt} + \frac{\partial A_i}{\partial t}.$$

But we can write $A_i = \partial(v^j A_j)/\partial v^i = \partial(\mathbf{v}\cdot\mathbf{A})/\partial v^i$ which gives us

$$[\mathbf{v}\times(\boldsymbol{\nabla}\times\mathbf{A})]_i = \frac{\partial}{\partial x^i}(\mathbf{v}\cdot\mathbf{A}) - \frac{d}{dt}\frac{\partial}{\partial v^i}(\mathbf{v}\cdot\mathbf{A}) + \frac{\partial A_i}{\partial t}.$$

The Lorentz force law can now be written in the form

$$\begin{aligned}
F_i &= e\left(-\frac{\partial\phi}{\partial x^i} - \frac{\partial A_i}{\partial t} + [\mathbf{v}\times(\boldsymbol{\nabla}\times\mathbf{A})]_i\right)\\
&= e\left(-\frac{\partial\phi}{\partial x^i} - \frac{\partial A_i}{\partial t} + \frac{\partial}{\partial x^i}(\mathbf{v}\cdot\mathbf{A}) - \frac{d}{dt}\frac{\partial}{\partial v^i}(\mathbf{v}\cdot\mathbf{A}) + \frac{\partial A_i}{\partial t}\right)\\
&= e\left[-\frac{\partial}{\partial x^i}(\phi - \mathbf{v}\cdot\mathbf{A}) - \frac{d}{dt}\frac{\partial}{\partial v^i}(\mathbf{v}\cdot\mathbf{A})\right].
\end{aligned}$$

Since ϕ is independent of $\mathbf{v}$ we can write

$$-\frac{d}{dt}\frac{\partial}{\partial v^i}(\mathbf{v}\cdot\mathbf{A}) = \frac{d}{dt}\frac{\partial}{\partial v^i}(\phi - \mathbf{v}\cdot\mathbf{A})$$

so that

$$F_i = e\left[-\frac{\partial}{\partial x^i}(\phi - \mathbf{v}\cdot\mathbf{A}) + \frac{d}{dt}\frac{\partial}{\partial v^i}(\phi - \mathbf{v}\cdot\mathbf{A})\right]$$

or

$$F_i = -\frac{\partial U}{\partial x_i} + \frac{d}{dt}\frac{\partial U}{\partial \dot{x}_i}$$

where $U = e(\phi - \mathbf{v}\cdot\mathbf{A})$. This shows that U is a generalized potential and that the Lagrangian for a particle of charge e in an electromagnetic field is

$$L = T - e\phi + e\mathbf{v}\cdot\mathbf{A} \tag{17a}$$

or

$$L = \frac{1}{2}mv^2 - e\phi + e\mathbf{v}\cdot\mathbf{A}. \tag{17b}$$

Since the canonical momentum is defined as $p_i = \partial L/\partial\dot{q}^i = \partial L/\partial v^i$ we now see that

$$p_i = mv_i + eA_i$$

or

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A}.$$

Therefore, in the absence of the electromagnetic field the Hamiltonian is $H = \mathbf{p}^2/2m$ where $\mathbf{p} = m\mathbf{v}$, so if the field is present we must now write

$$H = \frac{(\mathbf{p} - e\mathbf{A})^2}{2m}. \tag{18}$$

(You can also think of this as writing $\mathbf{v} = (\mathbf{p} - e\mathbf{A})/m$ and now making this substitution in the definition $H = p^i\dot{q}_i - L$ so that $H = H(q,p)$.) Furthermore, the energy of the particle now has an additional term $-e\phi$ due to the work in moving against the $\mathbf{E}$ field, so the energy operator must be $E - e\phi$ (the magnetic field $\mathbf{B}$ does no work since the force is perpendicular to the velocity).

We can combine these results into a single relativistic 4-momentum by making the replacement

$$p^\mu \to p^\mu - eA^\mu \tag{19}$$

where the 4-potential is given by $A^\mu = (\phi, \mathbf{A})$. The Dirac Hamiltonian now becomes

$$H_D = \boldsymbol{\alpha} \cdot \mathbf{p} - e\boldsymbol{\alpha} \cdot \mathbf{A} + \beta m + e\phi + V \tag{20}$$

where V is any additional potential that may be acting on the particle. Making the replacement (19) in the case where $V = 0$, the Dirac equation becomes

$$(\not{p} - e\not{A} - m)\psi = 0. \tag{21}$$

One of the great triumphs of the Dirac equation is that it gives us the correct gyromagnetic ratio with $g = 2$ for the electron, and the correct form of the spin-orbit coupling including the Thomas factor of $1/2$. We are now in the position to prove these results.

To see all of this, first write the Dirac equation as two coupled spinor equations, and then take the non-relativistic limit. Using the Hamiltonian (20) and the standard representation for the Dirac matrices, the Dirac equation can be written in the form

$$i\partial_t \begin{bmatrix} \varphi \\ \chi \end{bmatrix} = \boldsymbol{\alpha} \cdot (\mathbf{p} - e\mathbf{A}) \begin{bmatrix} \varphi \\ \chi \end{bmatrix} + \beta m \begin{bmatrix} \varphi \\ \chi \end{bmatrix} + e\phi \begin{bmatrix} \varphi \\ \chi \end{bmatrix}$$

where $\mathbf{p} = -i\boldsymbol{\nabla}$ and $\boldsymbol{\alpha}$ and β are given by equations (5). This is equivalent to the coupled equations

$$i\partial_t \varphi = \boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\chi + (e\phi + m)\varphi \tag{22a}$$

$$i\partial_t \chi = \boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\varphi + (e\phi - m)\chi \tag{22b}$$

In the non-relativistic limit, m is the largest energy term in this equation, so we write

$$\begin{bmatrix} \varphi \\ \chi \end{bmatrix} = e^{-imt} \begin{bmatrix} \widetilde{\varphi} \\ \widetilde{\chi} \end{bmatrix}$$

where the 2-component spinors $\widetilde{\varphi}$ and $\widetilde{\chi}$ are relatively slowly varying functions of time.

Using (22b) we have

$$i\partial_t \widetilde{\chi} = \boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\widetilde{\varphi} + (e\varphi - 2m)\widetilde{\chi} \approx 0.$$

Since $e\varphi \ll 2m$ this becomes

$$\widetilde{\chi} = \frac{1}{2m}\boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\widetilde{\varphi}$$

and since $\mathbf{p} \approx m\mathbf{v}$, we see that $\widetilde{\chi} \sim O(v/c) \times \widetilde{\varphi}$. (Remember we are using units where $c = 1$.) Because of this, we refer to $\widetilde{\chi}$ as the "small component" and $\widetilde{\varphi}$ as the "large component" of ψ.

Substituting the above expression for $\widetilde{\chi}$ into (22a) we obtain

$$i\partial_t \widetilde{\varphi} = \frac{1}{2m}[\boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})]^2 \widetilde{\varphi} + e\phi\widetilde{\varphi}.$$

From equation (13b) we have the very useful result

$$(\boldsymbol{\sigma} \cdot \mathbf{a})(\boldsymbol{\sigma} \cdot \mathbf{b}) = \mathbf{a} \cdot \mathbf{b} + i\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b}). \tag{23}$$

Writing $\boldsymbol{\pi} := \mathbf{p} - e\mathbf{A}$ and using (23) we have

$$(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2 = \boldsymbol{\pi} \cdot \boldsymbol{\pi} + i\boldsymbol{\sigma} \cdot (\boldsymbol{\pi} \times \boldsymbol{\pi}).$$

Note that $\boldsymbol{\pi}$ is a differential operator so that $\boldsymbol{\pi} \times \boldsymbol{\pi} \neq 0$. In particular, we have

$$\begin{aligned} \boldsymbol{\pi} \times \boldsymbol{\pi} &= (\mathbf{p} - e\mathbf{A}) \times (\mathbf{p} - e\mathbf{A}) = -e(\mathbf{A} \times \mathbf{p} + \mathbf{p} \times \mathbf{A}) \\ &= ie(\mathbf{A} \times \boldsymbol{\nabla} + \boldsymbol{\nabla} \times \mathbf{A}) \end{aligned}$$

so that

$$(\boldsymbol{\pi} \times \boldsymbol{\pi})\widetilde{\varphi} = ie[\mathbf{A} \times \boldsymbol{\nabla}\widetilde{\varphi} + \boldsymbol{\nabla} \times (\mathbf{A}\widetilde{\varphi})] = ie(\boldsymbol{\nabla} \times \mathbf{A})\widetilde{\varphi} = ie\mathbf{B}\widetilde{\varphi}.$$

Therefore we have

$$i\frac{\partial \widetilde{\varphi}}{\partial t} = \frac{(\mathbf{p} - e\mathbf{A})^2}{2m}\widetilde{\varphi} - \frac{e}{2m}(\boldsymbol{\sigma} \cdot \mathbf{B})\widetilde{\varphi} + e\phi\,\widetilde{\varphi}. \tag{24}$$

This is the non-relativistic Pauli equation for a particle of spin 1/2 in an electromagnetic field. Note that the magnetic moment is *predicted* to be $e/2m$ (in other units, this is $e\hbar/2mc$), and thus we automatically have $g = 2$ exactly. (There are higher order corrections to this that follow from the formalism of QED.)

(A sketch of the classical theory is as follows: The orbital magnetic moment of a current loop is

$$\mu_l = \frac{I}{c} \times \text{area}$$

where

$$I = \frac{\text{charge}}{\text{time}} = \frac{\text{charge}}{\text{dist/vel}} = \frac{e}{2\pi r/v} = \frac{ev}{2\pi r}$$

so that

$$\mu_l = \frac{ev}{2\pi rc}\pi r^2 = \frac{evr}{2c} = \frac{eL}{2mc}.$$

As vectors, this is

$$\boldsymbol{\mu}_l = \frac{e}{2mc}\mathbf{L}$$

where the ratio of $\boldsymbol{\mu}$ to $\mathbf{L}$ is called the **gyromagnetic ratio** γ.

Generalizing this, we make the definition

$$\boldsymbol{\mu} = g\frac{e}{2mc}\mathbf{J}$$

where g is a constant. For an electron we have $\mathbf{J} = \mathbf{S} = (\hbar/2)\boldsymbol{\sigma}$ and, from experiment, g is very close to 2 so that

$$\boldsymbol{\mu}_s = \frac{e\hbar}{2mc}\boldsymbol{\sigma}$$

with an energy $-\boldsymbol{\mu}_s \cdot \mathbf{B} = -(e\hbar/2mc)(\boldsymbol{\sigma} \cdot \mathbf{B})$.)

Now for the spin-orbit coupling in a hydrogen atom (with a nucleus of essentially infinite mass). To describe this, we first rewrite equations (22) by looking for energy eigenstates $\psi(\mathbf{x}, t) = e^{-iEt}\psi(\mathbf{x})$. Then we can write equations (22) as

$$E\varphi = \boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\chi + (e\phi + m)\varphi \tag{25a}$$

$$E\chi = \boldsymbol{\sigma} \cdot (\mathbf{p} - e\mathbf{A})\varphi + (e\phi - m)\chi \tag{25b}$$

where now φ and χ are independent of time. With $\mathbf{A} = 0$ (there is no external field) and letting $e\phi = V$, the second of these may be written as

$$\chi = (E - V + m)^{-1}(\boldsymbol{\sigma} \cdot \mathbf{p})\varphi.$$

(Be sure to remember that $\mathbf{p} = -i\boldsymbol{\nabla}$ so the order of factors is important because V is not a constant.) Let $E = E' + m$ so that

$$\chi = (E' - V + 2m)^{-1}(\boldsymbol{\sigma} \cdot \mathbf{p})\varphi.$$

Putting this into (25a) we can write

$$E'\varphi = \frac{(\boldsymbol{\sigma} \cdot \mathbf{p})}{2m}\left(1 + \frac{E' - V}{2m}\right)^{-1}(\boldsymbol{\sigma} \cdot \mathbf{p})\varphi + V\varphi$$

and to first order this is

$$E'\varphi = \frac{(\boldsymbol{\sigma} \cdot \mathbf{p})}{2m}\left(1 - \frac{E' - V}{2m}\right)(\boldsymbol{\sigma} \cdot \mathbf{p})\varphi + V\varphi.$$

From $[\mathbf{p}, V] = -i\boldsymbol{\nabla}V$ we have $\mathbf{p}V = V\mathbf{p} - i\boldsymbol{\nabla}V$ so our equation becomes

$$E'\varphi = \left[\frac{1}{2m}\left(1 - \frac{E' - V}{2m}\right)(\boldsymbol{\sigma} \cdot \mathbf{p})^2 - \frac{i}{4m^2}(\boldsymbol{\sigma} \cdot \boldsymbol{\nabla}V)(\boldsymbol{\sigma} \cdot \mathbf{p})\right]\varphi + V\varphi.$$

Using (23) we have $(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = p^2$ and

$$(\boldsymbol{\sigma} \cdot \boldsymbol{\nabla}V)(\boldsymbol{\sigma} \cdot \mathbf{p}) = \boldsymbol{\nabla}V \cdot \mathbf{p} - i\boldsymbol{\sigma} \cdot (\boldsymbol{\nabla}V \times \mathbf{p})$$

so that

$$E'\varphi = \left[\left(1 - \frac{E' - V}{2m}\right)\frac{p^2}{2m} - \frac{i}{4m^2}\boldsymbol{\nabla}V \cdot \mathbf{p} - \frac{1}{4m^2}\boldsymbol{\sigma} \cdot (\boldsymbol{\nabla}V \times \mathbf{p})\right]\varphi + V\varphi.$$

We assume spherical symmetry for V (for the hydrogen atom $V = -e^2/r$) so that

$$\boldsymbol{\nabla} V = \frac{dV}{dr}\hat{\mathbf{r}} \implies \boldsymbol{\nabla} V \cdot \mathbf{p} = -i\frac{dV}{dr}\frac{\partial}{\partial r}$$

and

$$\boldsymbol{\nabla} V \times \mathbf{p} = \frac{dV}{dr}\hat{\mathbf{r}} \times \mathbf{p} = \frac{1}{r}\frac{dV}{dr}\mathbf{r} \times \mathbf{p} = \frac{1}{r}\frac{dV}{dr}\mathbf{L}.$$

Therefore, since $\mathbf{S} = \boldsymbol{\sigma}/2$ we have

$$E'\varphi = \left[\left(1 - \frac{E' - V}{2m}\right)\frac{p^2}{2m} + V\right]\varphi - \frac{1}{4m^2}\frac{dV}{dr}\frac{\partial \varphi}{\partial r} - \frac{1}{2m^2}\frac{1}{r}\frac{dV}{dr}(\mathbf{S}\cdot\mathbf{L})\varphi.$$

Finally, since E is the total energy, we can write $E' - V \approx p^2/2m$ to arrive at the Schrödinger-like (two-component) equation

$$\left[\frac{p^2}{2m} - \frac{p^4}{8m^3} + V - \frac{1}{4m^2}\frac{dV}{dr}\frac{\partial}{\partial r} - \frac{1}{2m^2}\frac{1}{r}\frac{dV}{dr}(\mathbf{S}\cdot\mathbf{L})\right]\varphi = E'\varphi. \tag{26}$$

The second term is a relativistic correction to the kinetic energy, and the fourth term is called the "Darwin term." It is essentially due to the fact that a relativistic particle can't be localized to within better than its Compton wavelength $\hbar/mc$, and as a result the effective potential is really smeared out. And lastly, the final term is the spin-orbit coupling including the factor of 1/2 from Thomas precession. (Very roughly, here is the non-relativistic approach: The electron sees a current due to the relative motion of the nucleus, and this is the source of a magnetic field

$$\mathbf{B} = (-e/c)\mathbf{v} \times \mathbf{r}/r^3 = (e/mc)\mathbf{p} \times \mathbf{r}/r^3 = (-e/mcr^3)\mathbf{L}.$$

Then there will be an interaction energy term in the Hamiltonian that is

$$-\boldsymbol{\mu}\cdot\mathbf{B} = -(e/mc)(-e/mcr^3)\mathbf{S}\cdot\mathbf{L} = (e^2/m^2c^2r^3)\mathbf{S}\cdot\mathbf{L}.$$

With $c = 1$ and $V = -e^2/r$, this is the same as $(1/m^2r)(dV/dr)(\mathbf{S}\cdot\mathbf{L})$. However, this answer is off by a factor of 1/2 due to Thomas precession, and this is automatically taken into account in equation (26).)

3 Covariance of the Dirac Equation

We now turn our attention to the covariance of the Dirac equation under a Lorentz transformation

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu. \tag{27}$$

Note that $x'^\mu x'_\mu = \Lambda^\mu{}_\nu \Lambda_\mu{}^\alpha x^\nu x_\alpha := x^\nu x_\nu$ which implies

$$\Lambda^\mu{}_\nu \Lambda_\mu{}^\alpha = {\Lambda^T}^\alpha{}_\mu \Lambda^\mu{}_\nu = \delta^\alpha_\nu = g^\alpha_\nu. \tag{28}$$

This shows that ${\Lambda^T}^\alpha{}_\mu = {\Lambda^{-1}}^\alpha{}_\mu$ so that Λ is an orthogonal transformation, i.e.,

$$(\Lambda^{-1})_{\alpha\mu} = (\Lambda^T)_{\alpha\mu} = \Lambda_{\mu\alpha}.$$

Equation (28) can also be written as $(\Lambda^T)_{\alpha\mu}\Lambda^\mu{}_\nu = (\Lambda^T)_{\alpha\mu}\, g^{\mu\beta}\Lambda_{\beta\nu} = g_{\alpha\nu}$ or most simply as

$$\Lambda^T g\Lambda = g \tag{29}$$

which is frequently taken as the *definition* of a Lorentz transformation Λ. Note in particular that since $\Lambda^T = \Lambda^{-1}$ we also have $\Lambda g \Lambda^T = g$ and therefore

$$g_{\alpha\beta} = g_{\mu\nu}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta = \Lambda_{\nu\alpha}\Lambda^\nu{}_\beta = \Lambda_\alpha{}^\nu\Lambda_{\beta\nu}. \tag{30}$$

Since ∂_μ is a 4-vector we have $\partial'_\mu = \Lambda_\mu{}^\nu\partial_\nu$, and inverting this yields $\partial_\alpha = (\Lambda^{-1})_\alpha{}^\mu\partial'_\mu = \Lambda^\mu{}_\alpha\partial'_\mu$. (That ∂_μ is a true 4-vector follows from equation (27). We first have $x^\alpha = (\Lambda^{-1})^\alpha{}_\mu x'^\mu$ so that $\partial x^\alpha/\partial x'^\mu = (\Lambda^{-1})^\alpha{}_\mu = \Lambda_\mu{}^\alpha$. Therefore

$$\partial'_\mu = \frac{\partial}{\partial x'^\mu} = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial}{\partial x^\alpha} = \Lambda_\mu{}^\alpha\frac{\partial}{\partial x^\alpha} = \Lambda_\mu{}^\alpha\partial_\alpha$$

which shows that ∂_μ indeed has the correct transformation properties.) Applying this to the Dirac equation we have

$$0 = (i\gamma^\mu\partial_\mu - m)\psi(x) = (i\gamma^\mu\Lambda^\alpha{}_\mu\partial'_\alpha - m)\psi(\Lambda^{-1}x')$$

Let us define $\gamma'^\alpha = \Lambda^\alpha{}_\mu\gamma^\mu$ and observe that (using equation (30))

$$\{\gamma'^\alpha, \gamma'^\beta\} = \Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu\{\gamma^\mu, \gamma^\nu\} = 2\Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu g^{\mu\nu} = 2\Lambda^\alpha{}_\mu\Lambda^{\beta\mu} = 2g^{\alpha\beta}$$

so the γ'^μ also obey equation (4). As we will prove below, **Pauli's Fundamental Theorem** shows that given any two sets of matrices $\{\gamma^\mu\}$ and $\{\gamma'^\mu\}$ satisfying the Clifford algebra (4), there exists a nonsingular matrix S such that

$$\gamma'^\alpha = \Lambda^\alpha{}_\mu\gamma^\mu = S^{-1}\gamma^\alpha S. \tag{31}$$

(That $\gamma'^\alpha = \Lambda^\alpha{}_\mu\gamma^\mu$ is simply our definition of γ'^α — it has nothing to do with the general conclusion of Pauli's theorem.)

We now use this result to write

$$\begin{aligned}0 &= (i\gamma^\mu\Lambda^\alpha{}_\mu\partial'_\alpha - m)\psi(\Lambda^{-1}x') = (iS^{-1}\gamma^\alpha S\partial'_\alpha - m)\psi(\Lambda^{-1}x')\\ &= (iS^{-1}\gamma^\alpha S\partial'_\alpha - S^{-1}Sm)\psi(\Lambda^{-1}x')\\ &= S^{-1}(i\gamma^\alpha\partial'_\alpha - m)S\psi(\Lambda^{-1}x')\end{aligned}$$

which then implies

$$0 = (i\partial\!\!\!/' - m)\psi'(x')$$

where we have defined the transformed wave function

$$\psi'(x') := S(\Lambda)\psi(\Lambda^{-1}x') = S(\Lambda)\psi(x). \tag{32}$$

It is important to realize that *the gamma matrices themselves do not change under a Lorentz transformation*. Everything will be fine if we can show that the transformed wave function $\psi'(x')$ has the correct physical interpretation in the primed frame, i.e., we want to show that $j^\mu \to j'^\mu = \Lambda^\mu{}_\nu j^\nu$. Before doing this however, we first go back and prove Pauli's fundamental theorem because we will need some of the results that we prove along the way.

First of all, we want 16 linearly independent 4×4 matrices. Since $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$, it follows that $(\gamma^\mu)^2 = \pm 1$, and hence we need only consider products of *distinct* gamma matrices. Note that from the binomial theorem, the number of combinations of n objects taken one at a time, two at a time, $\dots$, n at a time is

$$\begin{aligned}\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = \sum_{k=1}^{n}\binom{n}{k} &= \sum_{k=0}^{n}\binom{n}{k}1^k 1^{n-k} - 1 \\ &= (1+1)^n - 1 = 2^n - 1.\end{aligned}$$

In our case we have $n = 4$, so there are 15 possible distinct combinations of the gamma matrices taken one, two, three and four at a time. Together with the identity matrix, this gives us the 16 matrices Γ_i defined by

$$\begin{array}{cc}
I & \Gamma_1 \\
\gamma^0 \quad i\gamma^1 \quad i\gamma^2 \quad i\gamma^3 & \Gamma_2 - \Gamma_5 \\
\gamma^0\gamma^1 \quad \gamma^0\gamma^2 \quad \gamma^0\gamma^3 \quad \gamma^1\gamma^2 \quad \gamma^1\gamma^3 \quad \gamma^2\gamma^3 & \Gamma_6 - \Gamma_{11} \\
i\gamma^0\gamma^1\gamma^2 \quad i\gamma^0\gamma^1\gamma^3 \quad i\gamma^0\gamma^2\gamma^3 \quad i\gamma^1\gamma^2\gamma^3 & \Gamma_{12} - \Gamma_{15} \\
i\gamma^0\gamma^1\gamma^2\gamma^3 & \Gamma_{16}
\end{array}$$

The factors of i are included so that

$$(\Gamma_i)^2 = +1. \tag{33}$$

Using the fact that the gamma matrices anticommute, it is easy to see that $\Gamma_i\Gamma_j = \pm\Gamma_j\Gamma_i$, and in fact

$$\Gamma_i\Gamma_j = a_{ij}\Gamma_k \qquad \text{where } a_{ij} = \pm 1, \pm i. \tag{34}$$

If $\Gamma_j \neq \Gamma_1$, there exists at least one Γ_i such that

$$\Gamma_i\Gamma_j\Gamma_i = -\Gamma_j. \tag{35}$$

In particular, we have

$$\begin{array}{lll}
\Gamma_j, 2 \le j \le 5 & \Longrightarrow & \Gamma_i = \Gamma_{16} \\
\Gamma_j, 6 \le j \le 11 & \Longrightarrow & \Gamma_i = \text{ whichever of the } \Gamma_2, \Gamma_3, \Gamma_4 \text{ or } \Gamma_5 \text{ that} \\
 & & \quad \text{contains one of the same } \gamma^\mu\text{'s that is in } \Gamma_j \\
\Gamma_j, 12 \le j \le 15 & \Longrightarrow & \Gamma_i = \Gamma_{16} \\
\Gamma_j = \Gamma_{16} & \Longrightarrow & \Gamma_i = \Gamma_2, \Gamma_3, \Gamma_4 \text{ or } \Gamma_5 \text{ all work}
\end{array} \tag{36}$$

Note that equations (33) and (35) together imply

$$\operatorname{tr}\Gamma_j = 0 \qquad \text{for } j \neq 1. \tag{37}$$

We still have to show that the Γ_i's are linearly independent. There are (at least) two ways to show this. First, suppose that $x_1\Gamma_1 + \cdots + x_{16}\Gamma_{16} = 0$. Taking the trace shows that $x_1 = 0$ since $\operatorname{tr}\Gamma_1 = 4 \neq 0$ and $\operatorname{tr}\Gamma_j = 0$ for $j \neq 1$. From (35) we have $\Gamma_i\Gamma_j = -\Gamma_j\Gamma_i$ (for $j \neq 1$), so it follows that $\operatorname{tr}\Gamma_i\Gamma_j = 0$ as long as $i \neq j$. Therefore, multiplying $x_2\Gamma_2 + \cdots + x_{16}\Gamma_{16} = 0$ by Γ_i and taking the trace implies that $x_i = 0$ for each $i = 2, \ldots, 16$. Therefore the Γ's are linearly independent and form a basis for the space of 4×4 complex matrices.

The second way to see this is to also start from $\sum_{k=1}^{16} x_k\Gamma_k = 0$. Multiplying by Γ_m we obtain

$$0 = x_m I + \sum_{k\neq m} x_k\Gamma_k\Gamma_m = x_m I + \sum_{k\neq m} x_k a_{km}\Gamma_n$$

where $\Gamma_n \neq I$ since $k \neq m$. (If $k \neq m$ and $\Gamma_k\Gamma_m = a_{km}I$, then $\Gamma_k = a_{km}\Gamma_m$ which is impossible.) Taking the trace now shows that $x_m = 0$.

In either case, we see that any 4×4 complex matrix X has a unique expansion $X = \sum x_i\Gamma_i$, since if we also have $X = \sum y_i\Gamma_i$, then $\sum(x_i - y_i)\Gamma_i = 0$ which implies that $x_i = y_i$ since the Γ's are linearly independent. Note also that the expansion coefficients x_i are determined by

$$\operatorname{tr}(X\Gamma_j) = x_j \operatorname{tr} I = 4x_j$$

or

$$x_i = \frac{1}{4}\operatorname{tr}(X\Gamma_i). \tag{38}$$

It is also true that $\Gamma_i\Gamma_j = a_{ij}\Gamma_k$ where Γ_k is different for each j (and fixed i). To see this, suppose $\Gamma_i\Gamma_j = a_{ij}\Gamma_k$ and $\Gamma_i\Gamma_{j'} = a_{ij'}\Gamma_k$. Multiplying from the left by Γ_i shows that $\Gamma_j = a_{ij}\Gamma_i\Gamma_k$ and $\Gamma_{j'} = a_{ij'}\Gamma_i\Gamma_k$ which implies $(1/a_{ij})\Gamma_j = (1/a_{ij'})\Gamma_{j'}$ or $\Gamma_j = (a_{ij}/a_{ij'})\Gamma_{j'}$ which contradicts the linear independence of the Γ's if $j \neq j'$.

The following theorem is sometimes called **Schur's lemma**, but technically that designation refers to irreducible group representations. In this case, the sixteen matrices Γ_i form a basis for what is called the **Dirac algebra**, which is a particular type of non-commutative ring. It can be shown that the only irreducible representation of the Dirac algebra is four-dimensional, but to do so would lead us too far astray from our present purposes.

Theorem 1. *If $X \in M_4(\mathbb{C})$ and $[X, \gamma^\mu] = 0$ for all μ, then $X = cI$ for some scalar c.*

Proof. Assume $X \neq cI$ and write $X = x_k\Gamma_k + \sum_{j\neq k} x_j\Gamma_j$ for any $k \neq 1$. From (35), there exists Γ_i such that $\Gamma_i\Gamma_k\Gamma_i = -\Gamma_k$. But $[X, \gamma^\mu] = 0$ implies

$[X, \Gamma_i] = 0$, and hence

$$X = \Gamma_i X \Gamma_i = x_k \Gamma_i \Gamma_k \Gamma_i + \sum_{j \neq k} x_j \Gamma_i \Gamma_j \Gamma_i = -x_k \Gamma_k + \sum_{j \neq k} \pm x_j \Gamma_j.$$

But the uniqueness of the expansion for X implies that $x_k = 0$. Since k was arbitrary except that $k \neq 1$, it follows that $X = x_1 \Gamma_1 \equiv cI$ (where we could have $c = 0$ if $X = 0$). ∎

We are now in a position to prove Pauli's theorem.

Theorem 2 (**Pauli's Fundamental Theorem**). *If* $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} = \{\overline{\gamma}^\mu, \overline{\gamma}^\nu\}$*, then there exists a nonsingular* S *such that* $\overline{\gamma}^\mu = S\gamma^\mu S^{-1}$*, and* S *is unique up to a multiplicative constant. (And hence we can always choose* $\det S = +1$*.)*

Proof. Define

$$S = \sum_{i=1}^{16} \overline{\Gamma}_i M \Gamma_i$$

where the $\overline{\Gamma}$'s are constructed from the $\overline{\gamma}$'s in exactly the same manner as the Γ's are from the γ's, and M is arbitrary. Note that M can always be chosen so that $S \neq 0$. Indeed, let $M_{rs} = \delta_{rr'}\delta_{ss'}$ have all 0 entries except for $M_{r's'} = 1$. Then

$$S_{pq} = \sum_{irs} (\overline{\Gamma}_i)_{pr} M_{rs} (\Gamma_i)_{sq} = \sum_i (\overline{\Gamma}_i)_{pr'} (\Gamma_i)_{s'q}.$$

If $S = 0$ for *all* M, then $S_{pq} = 0$ for all p, q and all s'. But then we have $0 = \sum_i (\overline{\Gamma}_i)_{pr'} \Gamma_i$ as a matrix equation (just take the $s'q$ entry of this equation), which contradicts the fact that the Γ's are linearly independent. Therefore, there exists M such that $S \neq 0$.

Now, $\Gamma_i \Gamma_j = a_{ij} \Gamma_k$ implies $\Gamma_i \Gamma_j \Gamma_i \Gamma_j = (a_{ij})^2 (\Gamma_k)^2 = (a_{ij})^2$, and hence multiplying by Γ_i from the left and Γ_j from the right yields

$$\Gamma_j \Gamma_i = (a_{ij})^2 \Gamma_i \Gamma_j = (a_{ij})^3 \Gamma_k.$$

Similarly, by definition it also follows that $\overline{\Gamma}_j \overline{\Gamma}_i = (a_{ij})^3 \overline{\Gamma}_k$. Using the fact that $(a_{ij})^4 = 1$ along with our earlier result that $\Gamma_i \Gamma_j = a_{ij} \Gamma_k$ where distinct Γ_j's correspond to distinct Γ_k's, we have

$$\overline{\Gamma}_i S \Gamma_i = \sum_j \overline{\Gamma}_i \overline{\Gamma}_j M \Gamma_j \Gamma_i = \sum_j a_{ij} \overline{\Gamma}_k M (a_{ij})^3 \Gamma_k = \sum_k \overline{\Gamma}_k M \Gamma_k = S$$

and therefore $S\Gamma_i = \overline{\Gamma}_i S$ or

$$\overline{\Gamma}_i = A \Gamma_i S^{-1}$$

if S^{-1} exists.

Defining $\overline{S} = \sum_i \Gamma_i \overline{M} \Gamma_i$ for arbitrary $\overline{M}$ yields (by symmetry with our previous result) $\Gamma_i \overline{S} \Gamma_i = \overline{S}$. Hence

$$\overline{S} S = \Gamma_i \overline{S} \overline{\Gamma}_i \overline{\Gamma}_i S \Gamma_i = \Gamma_i \overline{S} S \Gamma_i$$

or $[\overline{S}S, \Gamma_i] = 0$, and therefore $\overline{S}S = cI$ by Schur's lemma. Since $S, \overline{S} \neq 0$ we have $S^{-1} = (1/c)\overline{S}$ and S is nonsingular. To prove uniqueness, suppose $S_1 \gamma^\mu S_1^{-1} = S_2 \gamma^\mu S_2^{-1}$. Then $S_2^{-1} S_1 \gamma^\mu = \gamma^\mu S_2^{-1} S_1$ which (by Schur's lemma) implies $S_2^{-1} S_1 = aI$ or $S_1 = aS_2$. ∎

We now return to showing that the transformed wave function $\psi'(x')$ has the correct physical interpretation in the primed frame, i.e., that $j'^\mu = \Lambda^\mu{}_\nu j^\nu$. Using equations (10) and (31), the fact that $\Lambda^\alpha{}_\mu$ is just a real number and the fact that $(\gamma^0)^{-1} = \gamma^0$ we have

$$\begin{aligned}
\Lambda^\alpha{}_\mu \gamma^\mu &= \Lambda^\alpha{}_\mu \gamma^0 \gamma^{\mu\dagger} \gamma^0 = \gamma^0 (\Lambda^\alpha{}_\mu \gamma^\mu)^\dagger \gamma^0 = \gamma^0 (S^{-1} \gamma^\alpha S)^\dagger \gamma^0 \\
&= \gamma^0 S^\dagger \gamma^{\alpha\dagger} S^{-1\dagger} \gamma^0 = (\gamma^0 S^\dagger \gamma^0) \gamma^\alpha (\gamma^0 S^{\dagger -1} \gamma^0) \\
&= (\gamma^0 S^\dagger \gamma^0) \gamma^\alpha (\gamma^0 S^\dagger \gamma^0)^{-1}.
\end{aligned}$$

But we also have $\Lambda^\alpha{}_\mu \gamma^\mu = S^{-1} \gamma^\alpha S$, so equating this with the above result shows that

$$\gamma^\alpha S (\gamma^0 S^\dagger \gamma^0) = S (\gamma^0 S^\dagger \gamma^0) \gamma^\alpha$$

and hence $S\gamma^0 S^\dagger \gamma^0$ commutes with γ^α. Applying Schur's lemma we have $S\gamma^0 S^\dagger \gamma^0 = cI$ or

$$S \gamma^0 S^\dagger = c \gamma^0. \tag{39}$$

Taking the adjoint of this equation shows that c is real. We set the normalization of S be requiring that $\det S = +1 = \det S^\dagger$, and hence taking the determinant of equation (39) shows that $c^4 = 1$ (since $\det(c\gamma^0) = c^4 \det \gamma^0$) so that $c = \pm 1$.

We now show that $c = +1$ if $\Lambda^0{}_0 > 0$, i.e., there is no time reversal. First multiplying equation (39) from the right by γ^0 and from the left by S^{-1} gives us $\gamma^0 S^\dagger \gamma^0 = cS^{-1}$, and therefore $S^\dagger \gamma^0 = c\gamma^0 S^{-1}$. We then have (using equation (31))

$$\begin{aligned}
S^\dagger S &= S^\dagger \gamma^0 \gamma^0 S = c \gamma^0 S^{-1} \gamma^0 S = c \gamma^0 \Lambda^0{}_\mu \gamma^\mu \\
&= c \gamma^0 \Lambda^0{}_0 \gamma^0 + c \gamma^0 \Lambda^0{}_i \gamma^i \\
&= c \Lambda^0{}_0 I + c \Lambda^0{}_i \gamma^0 \gamma^i.
\end{aligned}$$

Since $S^\dagger S$ is Hermitian, it's eigenvalues are real. Alternatively, if $(S^\dagger S)x = \lambda x$ where x is normalized to $\|x\| = 1$, then $\lambda = \langle x, S^\dagger S x \rangle = \langle Sx, Sx \rangle = \|Sx\|^2 > 0$. (That $\|Sx\| \neq 0$ follows because the norm is positive definite, and the fact that S is nonsingular means $Sx = 0$ if and only if $x = 0$ which can't be true by definition of eigenvector.) In any case, we have $\operatorname{tr} S^\dagger S = \sum \lambda_i > 0$, and since

$\gamma^0\gamma^i = -\gamma^i\gamma^0$ we see that $\operatorname{tr}\gamma^0\gamma^i = 0$. But then taking the trace of the above expression for $S^\dagger S$ we obtain

$$0 < \operatorname{tr} S^\dagger S = \operatorname{tr}(c\Lambda^0{}_0 I) = 4c\Lambda^0{}_0$$

and we conclude that

$$\Lambda^0{}_0 > 0 \implies c = +1$$

and

$$\Lambda^0{}_0 < 0 \implies c = -1$$

as claimed. Since we restrict ourselves to the so-called orthochronous Lorentz transformations with $\Lambda^0{}_0 > 0$ then $c = +1$, and we have $S\gamma^0 S^\dagger = \gamma^0$ or

$$S^\dagger\gamma^0 = \gamma^0 S^{-1}. \tag{40}$$

(As a side remark just for the sake of complete accuracy, it follows from equation (30) that

$$1 = g_{00} = g_{\mu\nu}\Lambda^\mu{}_0\Lambda^\nu{}_0 = (\Lambda^0{}_0)^2 - \sum_{i=1}^{3}(\Lambda^i{}_0)^2$$

and therefore $(\Lambda^0{}_0)^2 = 1 + \sum_i(\Lambda^i{}_0)^2 \geq 1$ so we actually have either $\Lambda^0{}_0 \geq 1$ or $\Lambda^0{}_0 \leq -1$.)

Back to the physics of the transformed wave function $\psi'(x') = S\psi(x)$. Taking the adjoint of this we have $\psi'^\dagger = \psi^\dagger S^\dagger$ so that using equation (40) we have

$$\overline{\psi'} = \psi'^\dagger\gamma^0 = \psi^\dagger S^\dagger\gamma^0 = \psi^\dagger\gamma^0 S^{-1} = \overline{\psi}S^{-1}. \tag{41}$$

Therefore

$$j'^\mu = \overline{\psi'}\gamma^\mu\psi' = \overline{\psi}S^{-1}\gamma^\mu S\psi = \Lambda^\mu{}_\alpha\overline{\psi}\gamma^\alpha\psi = \Lambda^\mu{}_\alpha j^\alpha$$

as desired. In other words, the probability current $j^\mu = \overline{\psi}\gamma^\mu\psi$ transforms as a 4-vector and validates our interpretation of

$$\psi'(x') = S(\Lambda)\psi(\Lambda^{-1}x') = S(\Lambda)\psi(x)$$

as the wavefunction as seen in the transformed frame. We will use this equation to write the arbitrary momentum free particle solutions of the Dirac equation in terms of the rest particle solutions (which are easy to derive).

In fact, it is the transformation law (31) together with equation (41) that gives us the various types of elementary particle properties described as scalar, pseudoscalar, vector and pseudovector. Let us take a more careful look at just what this means.

The equations of motion are determined by a Lagrangian density $\mathscr{L}$ which is always a Lorentz scalar. But the terms that comprise $\mathscr{L}$ can vary widely. For example, consider the "scalar" $\overline{\psi}\psi$. That this is indeed a Lorentz scalar follows by direct calculation:

$$\overline{\psi'}\psi' = \overline{\psi}S^{-1}S\psi = \overline{\psi}\psi.$$

Furthermore, we just showed in the calculation above that the quantity $\overline{\psi}\gamma^\mu\psi$ transforms as a true 4-vector. What about the pseudo quantities? To treat these, we introduce the extremely useful gamma matrix

$$\gamma^5 := \gamma_5 := i\gamma^0\gamma^1\gamma^2\gamma^3 = \frac{i}{4!}\varepsilon_{\alpha\beta\mu\nu}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu. \tag{42}$$

That this last equality is true follows from the fact that all four indices must be distinct or else the ε symbol vanishes, and the gamma matrices all anticommute. Thus there are 4! possible permutations of four distinct gamma matrices, and putting these into increasing order introduces the same sign as the ε symbol acquires so all terms have the coefficient $+1$.

Under a Lorentz transformation we have

$$\begin{aligned}
S^{-1}\gamma_5 S &= \frac{i}{4!}\varepsilon_{\alpha\beta\mu\nu}S^{-1}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu S \\
&= \frac{i}{4!}\varepsilon_{\alpha\beta\mu\nu}S^{-1}\gamma^\alpha SS^{-1}\gamma^\beta SS^{-1}\gamma^\mu SS^{-1}\gamma^\nu S \\
&= \frac{i}{4!}\varepsilon_{\alpha\beta\mu\nu}\Lambda^\alpha{}_{\alpha'}\Lambda^\beta{}_{\beta'}\Lambda^\mu{}_{\mu'}\Lambda^\nu{}_{\nu'}\gamma^{\alpha'}\gamma^{\beta'}\gamma^{\mu'}\gamma^{\nu'} \\
&= \frac{i}{4!}(\det\Lambda)\varepsilon_{\alpha'\beta'\mu'\nu'}\gamma^{\alpha'}\gamma^{\beta'}\gamma^{\mu'}\gamma^{\nu'}
\end{aligned}$$

or

$$S^{-1}\gamma_5 S = (\det\Lambda)\gamma_5 \tag{43}$$

which shows that γ_5 transforms as a pseudoscalar, i.e., it depends on the sign of $\det\Lambda = \pm 1$. (Compare this with equation (31) which shows that γ^μ transforms as a vector.)

Using this result we can easily show that $\overline{\psi}\gamma_5\psi$ transforms as a pseudoscalar and $\overline{\psi}\gamma_5\gamma^\mu\psi$ transforms as a pseudovector:

$$\overline{\psi'}\gamma_5\psi' = \overline{\psi}S^{-1}\gamma_5 S\psi = (\det\Lambda)\overline{\psi}\gamma_5\psi$$

and

$$\begin{aligned}
\overline{\psi'}\gamma_5\gamma^\mu\psi' = \overline{\psi}S^{-1}\gamma_5\gamma^\mu S\psi &= \overline{\psi}S^{-1}\gamma_5 SS^{-1}\gamma^\mu S\psi \\
&= (\det\Lambda)\Lambda^\mu{}_\nu(\overline{\psi}\gamma_5\gamma^\nu\psi).
\end{aligned}$$

4 Construction of the Matrix $S(\Lambda)$

We begin by considering an infinitesimal Lorentz transformation

$$\Lambda^\mu{}_\nu = g^\mu{}_\nu + \omega^\mu{}_\nu. \tag{44}$$

Then to first order in ω we have

$$g_{\alpha\beta} = \Lambda^\mu{}_\alpha\Lambda_{\mu\beta} = (g^\mu{}_\alpha + \omega^\mu{}_\alpha)(g_{\mu\beta} + \omega_{\mu\beta}) = g_{\alpha\beta} + \omega_{\alpha\beta} + \omega_{\beta\alpha}$$

and thus

$$\omega_{\alpha\beta} = -\omega_{\beta\alpha}.$$

Let us expand $S(\Lambda)$ to first order in the parameters $\omega_{\mu\nu}$ to write

$$S = 1 - \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu} \tag{45a}$$

$$S^{-1} = 1 + \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}. \tag{45b}$$

Note that $\omega_{\mu\nu}$ is a *number*, while $\Sigma^{\mu\nu}$ is a 4×4 *matrix*. Since $\omega_{\mu\nu} = -\omega_{\nu\mu}$ we can antisymmetrize over μ and ν so that $\omega_{\mu\nu}\Sigma^{\mu\nu} = \omega_{\mu\nu}\Sigma^{[\mu\nu]}$ and hence we may just as well assume that Σ is antisymmetric, i.e.,

$$\Sigma^{\mu\nu} = -\Sigma^{\nu\mu}.$$

Just to clarify the antisymmetrization of $\Sigma^{\mu\nu}$, note that in general if we have an antisymmetric quantity $A_{\mu\nu}$ contracted with an arbitary quantity $T^{\mu\nu}$, then we always have

$$\begin{aligned}
A_{\mu\nu}T^{\mu\nu} &= \frac{1}{2}(A_{\mu\nu}T^{\mu\nu} + A_{\mu\nu}T^{\mu\nu}) \\
&= \frac{1}{2}(A_{\mu\nu}T^{\mu\nu} - A_{\nu\mu}T^{\mu\nu}) \qquad \text{by the antisymmetry of } A_{\mu\nu} \\
&= \frac{1}{2}(A_{\mu\nu}T^{\mu\nu} - A_{\mu\nu}T^{\nu\mu}) \qquad \text{by relabeling } \mu \leftrightarrow \nu \\
&= \frac{1}{2}A_{\mu\nu}(T^{\mu\nu} - T^{\nu\mu}) \\
&= A_{\mu\nu}T^{[\mu\nu]}.
\end{aligned}$$

This is an extremely useful property that we will use often. Note also that the quantity T can have additional indices that don't enter into the antisymmetrization, e.g., $A_{\mu\nu}T^{\mu\nu\rho} = A_{\mu\nu}T^{[\mu\nu]\rho}$.

Working to first order, we substitute equations (44) and (45) into equation (31):

$$(g^{\alpha}{}_{\mu} + \omega^{\alpha}{}_{\mu})\gamma^{\mu} = \left(1 + \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}\right)\gamma^{\alpha}\left(1 - \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}\right)$$

or

$$\gamma^{\alpha} + \omega^{\alpha}{}_{\mu}\gamma^{\mu} = \gamma^{\alpha} - \frac{i}{2}\omega_{\mu\nu}\gamma^{\alpha}\Sigma^{\mu\nu} + \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}\gamma^{\alpha}$$

which implies that

$$\omega^{\alpha}{}_{\mu}\gamma^{\mu} = -\frac{i}{2}\omega_{\mu\nu}[\gamma^{\alpha}, \Sigma^{\mu\nu}].$$

On the right hand side of this equation $\omega_{\mu\nu}$ is contracted with an antisymmetric quantity, so we want to do the same on the left. To accomplish this, we rewrite the left hand side as

$$\omega^{\alpha}{}_{\mu}\gamma^{\mu} = \omega_{\beta\mu}g^{\alpha\beta}\gamma^{\mu} = \omega_{\beta\mu}g^{\alpha[\beta}\gamma^{\mu]} = \frac{1}{2}\omega_{\beta\mu}(g^{\alpha\beta}\gamma^{\mu} - g^{\alpha\mu}\gamma^{\beta})$$
$$= \frac{1}{2}\omega_{\mu\nu}(g^{\alpha\mu}\gamma^{\nu} - g^{\alpha\nu}\gamma^{\mu}).$$

Therefore, since $\omega_{\mu\nu}$ is arbitrary, we must in fact have (after multiplying through by i)

$$i(g^{\alpha\mu}\gamma^{\nu} - g^{\alpha\nu}\gamma^{\mu}) = [\gamma^{\alpha}, \Sigma^{\mu\nu}].$$

Now, $\Sigma^{\mu\nu}$ is antisymmetric and, as we have seen, it must be a linear combination of the Γ matrices, i.e., it must be a product of γ matrices. If $\mu \neq \nu$ we know that $\gamma^{\mu}\gamma^{\nu} = -\gamma^{\nu}\gamma^{\mu}$, and if $\mu = \nu$ then obviously $[\gamma^{\mu}, \gamma^{\nu}] = 0$. Hence we try something of the form $\Sigma^{\mu\nu} \sim [\gamma^{\mu}, \gamma^{\nu}]$, and it is reasonably straightforward to verify that

$$\Sigma^{\mu\nu} = \frac{i}{4}[\gamma^{\mu}, \gamma^{\nu}] \tag{46}$$

will work. (To verify this, you will find it useful to note that $\gamma^{\alpha}\gamma^{\nu} + \gamma^{\nu}\gamma^{\alpha} = 2g^{\alpha\nu}$ implies $[\gamma^{\alpha}, \gamma^{\nu}] = \gamma^{\alpha}\gamma^{\nu} - \gamma^{\nu}\gamma^{\alpha} = 2(\gamma^{\alpha}\gamma^{\nu} - g^{\alpha\nu})$ along with the general commutator identity $[a, bc] = b[a, c] + [a, b]c$.) Thus we finally obtain (for infinitesimal $\omega_{\mu\nu}$)

$$S(\Lambda) = 1 - \frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu} = 1 + \frac{1}{8}\omega_{\mu\nu}[\gamma^{\mu}, \gamma^{\nu}]. \tag{47}$$

We now turn our attention to constructing $S(\Lambda)$ for finite $\Lambda^{\mu}{}_{\nu}$. Since a finite transformation consists of a product of a (infinite) number of infinitesimal transformations, we first prove a very useful mathematical result that you may have seen used in an elementary quantum mechanics course to construct the finite rotation operators $U(R(\boldsymbol{\theta})) = e^{i\boldsymbol{\theta}\cdot\mathbf{J}/\hbar}$.

Lemma.

$$\lim_{n\to\infty}\left(1 + \frac{\theta}{n}\right)^{n} = e^{\theta}.$$

Proof. First note that the logarithm is a continuous function (i.e., $\lim_{x\to a} f(x) = f(a)$) so that

$$\ln\lim_{n\to\infty}\left(1 + \frac{\theta}{n}\right)^{n} = \lim_{n\to\infty}\ln\left(1 + \frac{\theta}{n}\right)^{n} = \lim_{n\to\infty} n\ln\left(1 + \frac{\theta}{n}\right)$$
$$= \lim_{n\to\infty}\frac{\ln\left(1 + \theta/n\right)}{1/n}.$$

As $n \to \infty$ both the numerator and denominator each go to zero, so we use l'Hôpital's rule and take the derivative of both with respect to n. This yields

$$\ln\lim_{n\to\infty}\left(1 + \frac{\theta}{n}\right)^{n} = \lim_{n\to\infty}\frac{\frac{-\theta/n^2}{1+\theta/n}}{-1/n^2} = \lim_{n\to\infty}\frac{\theta}{1 + \theta/n} = \theta.$$

Taking the exponential of both sides then proves the lemma. ∎

Now, equations (45) and (47) apply to an infinitesimal boost parameter $\omega_{\mu\nu}$. In the case of a finite boost, let us write (as $n \to \infty$)

$$\omega_{\mu\nu} = \frac{\omega}{n}\widehat{\omega}_{\mu\nu}$$

which is a product of a Lorentz boost parameter ω and a *unit* Lorentz transformation matrix $\widehat{\omega}_{\mu\nu}$ (to be defined carefully below). Then this finite Λ is comprised of an infinite number of infinitesimal boosts and we have

$$S(\Lambda) = \lim_{n\to\infty}\left(1 - \frac{i}{2}\frac{\omega}{n}\widehat{\omega}_{\mu\nu}\Sigma^{\mu\nu}\right)^n = e^{-\frac{i}{2}\omega\widehat{\omega}_{\mu\nu}\Sigma^{\mu\nu}} \tag{48}$$

Let us verify equation (40) for this S. We first recall that $\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0$ and therefore

$$\begin{aligned}\Sigma^{\mu\nu\dagger} &= \left(\frac{i}{4}[\gamma^\mu,\gamma^\nu]\right)^\dagger = -\frac{i}{4}[\gamma^{\nu\dagger},\gamma^{\mu\dagger}] = \frac{i}{4}[\gamma^{\mu\dagger},\gamma^{\nu\dagger}] = \gamma^0\frac{i}{4}[\gamma^\mu,\gamma^\nu]\gamma^0 \\ &= \gamma^0\Sigma^{\mu\nu}\gamma^0.\end{aligned}$$

Writing $\omega\widehat{\omega}_{\mu\nu} = \omega_{\mu\nu}$ we then have (using the fact that $(\gamma^0)^2 = I$ to bring the γ^0's out of the exponential)

$$S = e^{-\frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}} \implies S^\dagger = e^{\frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu\dagger}} = \gamma^0 e^{\frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}}\gamma^0 = \gamma^0 S^{-1}\gamma^0$$

so that again we find $\gamma^0 S^\dagger = S^{-1}\gamma^0$ or $S\gamma^0 S^\dagger = \gamma^0$.

We now wish to construct an explicit form for S. To accomplish this, we need to know the boost generators $\widehat{\omega}_{\mu\nu}$. We know that for a Lorentz transformation along the positive x-axis we have (where the "lab frame" is labeled by x^μ and the "moving frame" is labeled by x'^μ)

$$x'^0 = \gamma(x^0 - \beta x^1) \qquad x'^1 = \gamma(x^1 - \beta x^0) \qquad x'^2 = x^2 \qquad x'^3 = x^3 \tag{49}$$

where $\beta = v/c$ and $\gamma^2 = 1/(1-\beta^2)$. To describe a boost in an arbitrary direction we first decompose this one into its components parallel and orthogonal to the velocity to write (keeping the axes of our coordinate systems parallel)

$$x'^0 = \gamma(x^0 - \boldsymbol{\beta}\cdot\mathbf{x}) \qquad \mathbf{x}'_\parallel = \gamma(\mathbf{x}_\parallel - \boldsymbol{\beta}x^0) \qquad \mathbf{x}'_\perp = \mathbf{x}_\perp$$

Now expand $\mathbf{x}'$ as follows:

$$\begin{aligned}\mathbf{x}' &= \mathbf{x}'_\perp + \mathbf{x}'_\parallel = \mathbf{x}_\perp + \gamma(\mathbf{x}_\parallel - \boldsymbol{\beta}x^0) = \mathbf{x} - \mathbf{x}_\parallel + \gamma(\mathbf{x}_\parallel - \boldsymbol{\beta}x^0) \\ &= \mathbf{x} + (\gamma - 1)\mathbf{x}_\parallel - \gamma\boldsymbol{\beta}x^0.\end{aligned}$$

But

$$\mathbf{x}_\parallel = (\mathbf{x}\cdot\widehat{\boldsymbol{\beta}})\widehat{\boldsymbol{\beta}} = \frac{\mathbf{x}\cdot\boldsymbol{\beta}}{\beta^2}\boldsymbol{\beta}$$

and hence we have

$$\mathbf{x}' = \mathbf{x} + \frac{(\gamma - 1)}{\beta^2}(\mathbf{x} \cdot \boldsymbol{\beta})\boldsymbol{\beta} - \gamma \boldsymbol{\beta} x^0 \tag{50a}$$

$$x'^0 = \gamma(x^0 - \boldsymbol{\beta} \cdot \mathbf{x}). \tag{50b}$$

Comparing these equations with $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ we write out the matrix $(\Lambda^\mu{}_\nu)$:

$$(\Lambda^\mu{}_\nu) = \begin{bmatrix} \gamma & -\gamma\beta^1 & -\gamma\beta^2 & -\gamma\beta^3 \\ -\gamma\beta^1 & 1 + \frac{(\gamma-1)}{(\beta)^2}(\beta^1)^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^1\beta^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^1\beta^3 \\ -\gamma\beta^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^2\beta^1 & 1 + \frac{(\gamma-1)}{(\beta)^2}(\beta^2)^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^2\beta^3 \\ -\gamma\beta^3 & \frac{(\gamma-1)}{(\beta)^2}\beta^3\beta^1 & \frac{(\gamma-1)}{(\beta)^2}\beta^3\beta^2 & 1 + \frac{(\gamma-1)}{(\beta)^2}(\beta^3)^2 \end{bmatrix}. \tag{51}$$

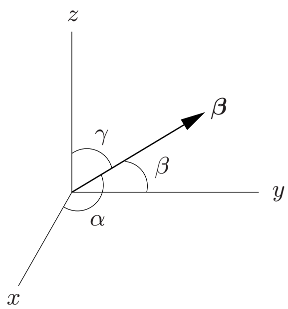
For an infinitesimal transformation we have $\gamma \to 1$ so that

$$(\Lambda^\mu{}_\nu) = \begin{bmatrix} 1 & -\beta^1 & -\beta^2 & -\beta^3 \\ -\beta^1 & 1 & & \\ -\beta^2 & & 1 & \\ -\beta^3 & & & 1 \end{bmatrix} = (g^\mu{}_\nu) + (\omega^\mu{}_\nu).$$

(Note that this is for a pure boost only. If we also included a spatial rotation, then the lower right 3×3 block would contain an infinitesimal rotation matrix.) In any case, we therefore have (in a somewhat ambiguous but standard notation)

$$\begin{aligned} (\omega^\mu{}_\nu) &= \begin{bmatrix} 0 & -\beta^1 & -\beta^2 & -\beta^3 \\ -\beta^1 & 0 & & \\ -\beta^2 & & 0 & \\ -\beta^3 & & & 0 \end{bmatrix} \\ &= \beta \begin{bmatrix} 0 & -\cos\alpha & -\cos\beta & -\cos\gamma \\ -\cos\alpha & 0 & & \\ -\cos\beta & & 0 & \\ -\cos\gamma & & & 0 \end{bmatrix} := \beta\,(\widehat{\omega}^\mu{}_\nu) \end{aligned} \tag{52}$$

which defines the unit transformation matrix $(\widehat{\omega}^\mu{}_\nu)$, and where $(\cos\alpha, \cos\beta, \cos\gamma)$ are the direction cosines of the infinitesimal boost $\beta = \omega/n$. (See the figure below.)



From the figure we have $\boldsymbol{\beta} = (\beta\cos\alpha)\hat{\mathbf{x}} + (\beta\cos\beta)\hat{\mathbf{y}} + (\beta\cos\gamma)\hat{\mathbf{z}}$ so $\beta^2 = \boldsymbol{\beta}\cdot\boldsymbol{\beta} = \beta^2(\cos^2\alpha + \cos^2\beta + \cos^2\gamma)$ and hence

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

Note also that we have defined the matrix $(\widehat{\omega}^\mu{}_\nu)$, and from this we can write $\widehat{\omega}_{\mu\nu} = g_{\mu\alpha}\,\widehat{\omega}^\alpha{}_\nu$. Then

$$\widehat{\omega}_{0i} = g_{0\alpha}\widehat{\omega}^\alpha{}_i = \widehat{\omega}^0{}_i \qquad \text{and} \qquad \widehat{\omega}_{i0} = g_{i\alpha}\widehat{\omega}^\alpha{}_0 = -\widehat{\omega}^i{}_0$$

so we see from (52) that indeed we have $\omega_{\mu\nu} = -\omega_{\nu\mu}$.

We can now use these $\widehat{\omega}_{\mu\nu}$'s to find another form of the Lorentz transformation matrix $\Lambda^\mu{}_\nu$ via exponentiation. Thus, a finite Lorentz transformation is now given by

$$\lim_{n\to\infty}\Big(g^\mu{}_\nu + \frac{\omega}{n}\widehat{\omega}^\mu{}_\nu\Big)^n = (e^{\omega\widehat{\omega}})^\mu{}_\nu = \Lambda^\mu{}_\nu \tag{53}$$

where ω is a finite boost. To explicitly evaluate this, we observe that

$$(\widehat{\omega}^\mu{}_\nu)^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2\alpha & \cos\alpha\cos\beta & \cos\alpha\cos\gamma \\ 0 & \cos\beta\cos\alpha & \cos^2\beta & \cos\beta\cos\gamma \\ 0 & \cos\gamma\cos\alpha & \cos\gamma\cos\beta & \cos^2\gamma \end{bmatrix}$$

and $(\widehat{\omega}^\mu{}_\nu)^3 = (\widehat{\omega}^\mu{}_\nu)$. We also note that

$$\cosh\theta = \frac{1}{2}(e^\theta + e^{-\theta}) = 1 + \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \cdots$$
$$\sinh\theta = \frac{1}{2}(e^\theta - e^{-\theta}) = \theta + \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots$$

and therefore

$$\begin{aligned}\Lambda^\mu{}_\nu = (e^{\omega\widehat{\omega}})^\mu{}_\nu &= \Big(1 + \omega\widehat{\omega} + \frac{\omega^2}{2!}\widehat{\omega}^2 + \frac{\omega^3}{3!}\widehat{\omega} + \frac{\omega^4}{4!}\widehat{\omega}^2 + \frac{\omega^5}{5!}\widehat{\omega} + \cdots\Big)^\mu{}_\nu \\ &= \Big[\widehat{\omega}\Big(\omega + \frac{\omega^3}{3!} + \frac{\omega^5}{5!} + \cdots\Big) + 1 - \widehat{\omega}^2 + \widehat{\omega}^2\Big(1 + \frac{\omega^2}{2!} + \frac{\omega^4}{4!} + \cdots\Big)\Big]^\mu{}_\nu \\ &= [1 - \widehat{\omega}^2 + \widehat{\omega}^2\cosh\omega + \widehat{\omega}\sinh\omega]^\mu{}_\nu.\end{aligned}$$

For example, in the particular case of a boost along the x^1-axis we have $\cos\alpha = 1$ and $\cos\beta = \cos\gamma = 0$ so that

$$\widehat{\omega}^\mu{}_\nu = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & & \\ 0 & & 0 & \\ 0 & & & 0 \end{bmatrix} \qquad\qquad (\widehat{\omega}^\mu{}_\nu)^2 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}$$

and hence

$$e^{\omega\widehat{\omega}} = \begin{bmatrix} \cosh\omega & -\sinh\omega & 0 & 0 \\ -\sinh\omega & \cosh\omega & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = (\Lambda^\mu{}_\nu).$$

Therefore, looking at the 0 component for example, we have

$$x'^0 = \Lambda^0{}_\nu x^\nu = x^0\cosh\omega - x^1\sinh\omega = \cosh\omega(x^0 - x^1\tanh\omega)$$

and comparing this with equation (49) shows that

$$\cosh\omega = \gamma \qquad \text{and} \qquad \tanh\omega = \beta \tag{54}$$

which should be familiar from more elementary courses. In other words, exponentiating an infinitesimal Lorentz boost gives back exactly the same transformation matrix as we could have written down directly from equation (49), which should have been expected.

Now let us finish computing the spinor transformation matrix $S(\Lambda)$ defined in equation (48). First, using $\widehat{\omega}^\mu{}_\nu$ as defined in equation (52) we have

$$\begin{aligned}\widehat{\omega}_{\mu\nu}\Sigma^{\mu\nu} &= g_{\mu\rho}\,\widehat{\omega}^\rho{}_\nu\Sigma^{\mu\nu} \\ &= \widehat{\omega}^0{}_1\Sigma^{01} + \widehat{\omega}^0{}_2\Sigma^{02} + \widehat{\omega}^0{}_3\Sigma^{03} - \widehat{\omega}^1{}_0\Sigma^{10} - \widehat{\omega}^2{}_0\Sigma^{20} - \widehat{\omega}^3{}_0\Sigma^{30}.\end{aligned}$$

But $\Sigma^{0i} = -\Sigma^{i0} = (i/4)[\gamma^0, \gamma^i]$ where the gamma matrices are given in equation (6) so that

$$\Sigma^{0i} = \frac{i}{2}\begin{bmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{bmatrix} := \frac{i}{2}\alpha^i.$$

Now observe that (as we saw above) $\widehat{\omega}^i{}_0 = g^{i\mu}\widehat{\omega}_{\mu 0} = -\widehat{\omega}_{i0} = \widehat{\omega}_{0i} = g_{0\mu}\widehat{\omega}^\mu{}_i = \widehat{\omega}^0{}_i$ (which also follows from the explicit form of equation (52)) and therefore

$$\begin{aligned}\widehat{\omega}_{\mu\nu}\Sigma^{\mu\nu} &= 2\,\widehat{\omega}^0{}_i\Sigma^{0i} = -2(\Sigma^{01}\cos\alpha + \Sigma^{02}\cos\beta + \Sigma^{03}\cos\gamma) \\ &= -i(\alpha^1\cos\alpha + \alpha^2\cos\beta + \alpha^3\cos\gamma) \\ &= -i\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}}.\end{aligned}$$

Recall that the Pauli matrices obey the relation (using the summation convention on repeated indices)

$$\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k$$

which implies

$$(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = \mathbf{a}\cdot\mathbf{b} + i(\mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}$$

so that $(\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\beta}})^2 = \widehat{\boldsymbol{\beta}}\cdot\widehat{\boldsymbol{\beta}} = 1$. Then from

$$\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 0 & \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\beta}} \\ \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\beta}} & 0 \end{bmatrix}$$

(remember this is a block matrix) we see that

$$(\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}})^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This then gives us

$$\begin{aligned}
S(\Lambda) = e^{-\frac{i}{2}\omega\widehat{\omega}_{\mu\nu}\Sigma^{\mu\nu}} &= e^{-\frac{1}{2}\omega\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}}} \\
&= I - \frac{\omega}{2}\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}} + \frac{1}{2!}\Big(\frac{\omega}{2}\Big)^2(\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}})^2 - \frac{1}{3!}\Big(\frac{\omega}{2}\Big)^3(\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}})^3 + \cdots \\
&= I - \boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}}\bigg[\frac{\omega}{2} + \frac{1}{3!}\Big(\frac{\omega}{2}\Big)^3 + \cdots\bigg] + I\bigg[\frac{1}{2!}\Big(\frac{\omega}{2}\Big)^2 + \frac{1}{4!}\Big(\frac{\omega}{2}\Big)^4 + \cdots\bigg]
\end{aligned}$$

or simply

$$S(\Lambda) = I\cosh\frac{\omega}{2} - (\boldsymbol{\alpha}\cdot\widehat{\boldsymbol{\beta}})\sinh\frac{\omega}{2}. \tag{55}$$

If this looks vaguely familiar to you, it's because you may recall from a quantum mechanics course that the rotation operator for spin 1/2 particles is given by

$$U(R(\boldsymbol{\theta})) = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}/\hbar} = e^{-i\boldsymbol{\theta}\cdot\boldsymbol{\sigma}/2\hbar} = I\cos\frac{\theta}{2} - i(\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\theta}})\sin\frac{\theta}{2}.$$

Anyway, to put equation (55) into a more useable form, we make note of the following identities:

$$\begin{aligned}
&\cosh^2 x - \sinh^2 x = 1 \\
&\sinh(x+y) = \sinh x\cosh y + \cosh x\sinh y \\
&\cosh(x+y) = \cosh x\cosh y + \sinh x\sinh y
\end{aligned}$$

and these then imply

$$\begin{aligned}
&1 + \cosh 2x = 2\cosh^2 x \\
&\cosh 2x - 1 = 2\sinh^2 x \\
&\cosh\frac{x}{2} = \bigg[\frac{1}{2}(1+\cosh x)\bigg]^{1/2} \\
&\sinh\frac{x}{2} = \bigg[\frac{1}{2}(\cosh x - 1)\bigg]^{1/2}.
\end{aligned}$$

Using equation (54) we then have

$$\cosh\frac{\omega}{2} = \left[\frac{1}{2}(1+\gamma)\right]^{1/2} \qquad \text{and} \qquad \sinh\frac{\omega}{2} = \left[\frac{1}{2}(\gamma-1)\right]^{1/2}.$$

Now we use the relativistic expressions $E = \gamma m$ and $\mathbf{p} = \gamma m \mathbf{v} = E\mathbf{v}$ so that $\widehat{\mathbf{p}} = \mathbf{p}/p = \mathbf{p}/Ev$. Since we want to boost *from* the rest frame of the particle *to* a frame where it has velocity $\mathbf{v}$, we have $\boldsymbol{\beta} = -\mathbf{v}$ and $\widehat{\boldsymbol{\beta}} = -\widehat{\mathbf{p}}$. We also have $E^2/m^2 = \gamma^2 = 1/(1-\beta^2)$ so that

$$\beta^2 = 1 - \frac{m^2}{E^2} = \frac{E^2 - m^2}{E^2} = \frac{(E+m)(E-m)}{E^2}$$

or

$$E\beta = [(E+m)(E-m)]^{1/2}.$$

Therefore

$$\left(\frac{\gamma+1}{2}\right)^{1/2} = \left(\frac{E/m+1}{2}\right)^{1/2} = \left(\frac{E+m}{2m}\right)^{1/2} = \cosh\frac{\omega}{2} \tag{56a}$$

$$\left(\frac{\gamma-1}{2}\right)^{1/2} = \left(\frac{E-m}{2m}\right)^{1/2} = \sinh\frac{\omega}{2} \tag{56b}$$

$$-\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\beta}} = +\boldsymbol{\sigma}\cdot\widehat{\mathbf{p}} = \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E\beta} = \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{[(E+m)(E-m)]^{1/2}} \tag{56c}$$

Using equations (56) in equation (55) then yields our desired final form

$$S(\Lambda) = \\ \left(\frac{E+m}{2m}\right)^{1/2} \begin{bmatrix} 1 & 0 & p_z/(E+m) & p_-/(E+m) \\ 0 & 1 & p_+/(E+m) & -p_z/(E+m) \\ p_z/(E+m) & p_-/(E+m) & 1 & 0 \\ p_+/(E+m) & -p_z/(E+m) & 0 & 1 \end{bmatrix} \tag{57}$$

where $p_\pm = p_x \pm ip_y$.

We now use equation (57) to write down the arbitrary momentum free particle solutions to the Dirac equation. For a particle at rest we have $\mathbf{p} = 0$, so the Dirac Hamiltonian (equation (2)) becomes simply $E = \boldsymbol{\alpha}\cdot\mathbf{p} + \beta m = \beta m$, and the Dirac equation is just

$$(i\gamma^0\partial_0 - m)\psi(x) = 0.$$

Using (in block form)

$$\gamma^0 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$$

we write the equation in the form

$$i\begin{bmatrix}1&&&\\&1&&\\&&-1&\\&&&-1\end{bmatrix}\partial_0\begin{bmatrix}\phi_1\\ \phi_2\\ \chi_1\\ \chi_2\end{bmatrix}=m\begin{bmatrix}\phi_1\\ \phi_2\\ \chi_1\\ \chi_2\end{bmatrix}.$$

The obvious solutions are of the form

$$\psi_r(x)=w_r(0)e^{-i\epsilon_r mt}$$

where

$$\epsilon_r=\begin{cases}+1 & \text{for } r=1,2\\ -1 & \text{for } r=3,4\end{cases}$$

and

$$w_1(0)=\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\qquad w_2(0)=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\qquad w_3(0)=\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\qquad w_4(0)=\begin{bmatrix}0\\0\\0\\1\end{bmatrix}.$$

Since $mt=\mathring{p}_0x^0=\mathring{p}_\mu x^\mu$ is a Lorentz scalar (where $^\circ$ means the rest frame), we may write the phase in the form $e^{-i\epsilon_r p_\mu x^\mu}$. And since the spinor part is given by $w_r(\mathbf{p})=S(\Lambda)w_r(0)$, the general solution is thus

$$\psi_r(x)=w_r(\mathbf{p})e^{-i\epsilon_r p_\mu x^\mu}=w_r(\mathbf{p})e^{-i\epsilon_r p\cdot x} \tag{58}$$

where the rth column of (57) gives $w_r(\mathbf{p})$. Recalling that $\overline{\psi}\psi$ is a Lorentz scalar, it is also easy to see directly from the columns of (57) that

$$\overline{w}_r(0)w_s(0)=\overline{w}_r(p)w_s(p)=\epsilon_r\delta_{rs}.$$

But note that we can multiply $w_r(0)$ by any constant to fix the normalization.

The last topic to cover in this section is to consider what happens under parity (i.e., space reflection). In this case equation (31) can not be solved by considering the infinitesimal transformation (44). Now we have the Lorentz transformation $t\to t$ and $\mathbf{x}\to-\mathbf{x}$ so the Lorentz matrix Λ_P is given by

$$(\Lambda_P)^\mu{}_\nu=\begin{bmatrix}1&&&\\&-1&&\\&&-1&\\&&&-1\end{bmatrix}=g_{\mu\nu}$$

and we seek a matrix denoted by P (rather than S) that satisfies

$$(\Lambda_P)^\mu{}_\nu\gamma^\nu=P^{-1}\gamma^\mu P.$$

In particular, $\gamma^0 = P^{-1}\gamma^0 P$ and $-\gamma^i = P^{-1}\gamma^i P$ or $-P\gamma^i = \gamma^i P$. It should be clear that this is satisfied by choosing anything of the form

$$P = e^{i\varphi}\gamma^0. \tag{59}$$

In other words, we have

$$\psi(x) \xrightarrow{P} \psi'(x') = \psi'(t, -\mathbf{x}) = e^{i\varphi}\gamma^0\psi(t, \mathbf{x}).$$

5 Easier Approach to the Spinor Solutions

The Dirac equation is $(\boldsymbol{\alpha}\cdot\mathbf{p} + \beta m)\psi = E\psi$, and with $\gamma^0 = \beta, \boldsymbol{\gamma} = \beta\boldsymbol{\alpha}$ we have $(\boldsymbol{\gamma}\cdot\mathbf{p} + m)\psi = \gamma^0 E\psi$. Using $p^\mu = (p^0 = p_0 = E, \mathbf{p})$ we write the Dirac equation as $(\gamma^0 p_0 - \boldsymbol{\gamma}\cdot\mathbf{p} - m)\psi = (\gamma^\mu p_\mu - m)\psi = 0$ or just

$$(\not{p} - m)\psi = 0. \tag{60}$$

(For simplicity we will generally leave out the identity matrix in these equations.)

Now note that multiplying $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ by scalars a_μ, b_ν we obtain

$$\{\not{a}, \not{b}\} = 2\, a\cdot b \tag{61}$$

and hence in particular $\not{p}\not{p} = p\cdot p = p^2$. Operating on equation (60) from the left with $\not{p} + m$ yields $(p^2 - m^2)\psi = 0$, and using $p^\mu = i\partial^\mu$ this becomes $(-\partial_\mu\partial^\mu - m^2)\psi = 0$ or

$$(\Box + m^2)\psi = 0.$$

In other words, each component of any ψ that satisfies the Dirac equation also satisfies the Klein-Gordon equation. This equation has the solutions $e^{\pm ik\cdot x}$ with $-k^2 + m^2 = 0$ so that $k_0^2 - \mathbf{k}^2 = m^2$ or $k_0^2 = \mathbf{k}^2 + m^2$. We define

$$\omega_k := +\sqrt{\mathbf{k}^2 + m^2} \tag{62}$$

so that the solution $e^{-ik\cdot x} \sim e^{-i\omega_k t}$ is referred to as the **positive frequency** solution (since in the Schrödinger theory $\psi \sim e^{-iEt}$), and the solution $e^{+ik\cdot x} \sim e^{i\omega_k t}$ is called the **negative frequency** solution.

Let us write the plane wave solutions to the Dirac equation in the form

$$\psi(x) \sim u(k)e^{-ik\cdot x} + v(k)e^{ik\cdot x}$$

where $k^2 = k_\mu k^\mu = m^2$. Then $(i\not{\partial} - m)\psi = 0$ implies

$$(+\not{k} - m)u(k)e^{-ik\cdot x} + (-\not{k} - m)v(k)e^{ik\cdot x} = 0.$$

Since the positive and negative frequency solutions are independent (they each satisfy the Dirac equation separately) this implies

$$(\not{k} - m)u(k) = 0 \tag{63a}$$

$$(\not{k} + m)v(k) = 0. \tag{63b}$$

Using $\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0$, we take the adjoint of each of these and multiply through by γ^0 to obtain

$$\overline{u}(k)(\not{k} - m) = 0 \tag{64a}$$

$$\overline{v}(k)(\not{k} + m) = 0. \tag{64b}$$

For solutions at rest we have $\mathbf{k} = 0$ (and hence $k_0 = m$) so equations (63) become

$$(\gamma^0 - 1)u(0) = 0$$

$$(\gamma^0 + 1)v(0) = 0.$$

The first of these is

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix} u(0) = u(0)$$

which has solutions of the form

$$u(0) = \begin{bmatrix} * \\ * \\ 0 \\ 0 \end{bmatrix}$$

where the *'s stand for an arbitrary entry. Similarly, the second of these has solutions of the form

$$v(0) = \begin{bmatrix} 0 \\ 0 \\ * \\ * \end{bmatrix}.$$

Since each of these has two independent components, we write the rest frame solutions as

$$u_1(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad u_2(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad v_1(0) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad v_2(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{65}$$

up to an arbitrary constant.

Using

$$(\not{k} - m)(\not{k} + m) = k^2 - m^2 \equiv 0 \tag{66}$$

along with the fact that $u(k)$ satisfies equation (63a), we see that any spinor of the form $u' = (\not{k} + m)u$ will automatically satisfy $(\not{k} - m)u' = 0$. We therefore write the solutions for arbitrary $\mathbf{k}$ in the form

$$u_r(k) = c(\not{k} + m)u_r(0) \tag{67a}$$

$$-v_r(k) = c'(\not{k} - m)v_r(0) \tag{67b}$$

where the normalization constants c and c' are to be determined, and the $(-)$ sign in front of $v_r(k)$ is an arbitrary convention. We must also therefore have (by inserting $\gamma^0\gamma^0 = I$ in front of the m)

$$u_r^\dagger(k) = c^* u_r^\dagger(0)(\gamma^0\gamma^\mu\gamma^0 k_\mu + m) = c^*\overline{u}_r(0)(\not{k} + m)\gamma^0$$

and

$$-v_r^\dagger(k) = c'^* v_r^\dagger(0)(\gamma^0\gamma^\mu\gamma^0 k_\mu - m) = c'^*\overline{v}_r(0)(\not{k} - m)\gamma^0$$

so that

$$\overline{u}_r(k) = c^*\overline{u}_r(0)(\not{k} + m) \tag{68a}$$

$$-\overline{v}_r(k) = c'^*\overline{v}_r(0)(\not{k} - m). \tag{68b}$$

Next, from equations (66), (67) and (68) we see that

$$\overline{u}_r(k)v_s(k) = -c^*c'\overline{u}_r(0)(\not{k} + m)(\not{k} - m)v_r(0) \equiv 0$$

and similarly

$$\overline{v}_r(k)u_s(k) = 0.$$

This means that $\overline{\psi}\psi \sim \overline{u}u + \overline{v}v$, and since u and v are independent solutions, it follows from the fact that $\overline{\psi}\psi$ is Lorentz invariant that $\overline{u}u$ and $\overline{v}v$ must also be Lorentz invariant. We fix our normalization by requiring that

$$\overline{u}_r(k)u_s(k) = \overline{u}_r(0)u_s(0) = 2m\delta_{rs} \tag{69a}$$

$$\overline{v}_r(k)v_s(k) = \overline{v}_r(0)v_s(0) = -2m\delta_{rs} \tag{69b}$$

where the $(-)$ sign in equation (69b) is due to the form of γ^0 and equations (65).

Let us now find the normalization constants in equations (67). We compute using equations (66), (67) and (68):

$$\begin{aligned}
\overline{u}_r(k)u_s(k) &= |c|^2\,\overline{u}_r(0)(\not{k} + m)^2 u_s(0) = |c|^2\,\overline{u}_r(0)(\not{k}^2 + 2m\not{k} + m^2)u_s(0) \\
&= |c|^2\,\overline{u}_r(0)2m(m + \not{k})u_s(0) = 2m\,|c|^2\,\overline{u}_r(0)(\not{k} + m)u_s(0) \\
&= 2m\,|c|^2\,u_r^\dagger(0)\gamma^0(\gamma^0 k_0 - \boldsymbol{\gamma}\cdot\mathbf{k} + m)u_s(0).
\end{aligned}$$

But

$$(\boldsymbol{\gamma}\cdot\mathbf{k})u_s(0) = \begin{bmatrix} 0 & \boldsymbol{\sigma}\cdot\mathbf{k} \\ -\boldsymbol{\sigma}\cdot\mathbf{k} & 0 \end{bmatrix}\begin{bmatrix} * \\ * \\ 0 \\ 0 \end{bmatrix} \sim \begin{bmatrix} 0 \\ 0 \\ * \\ * \end{bmatrix}$$

while

$$\overline{u}_r(0) = u_r^\dagger(0)\gamma^0 = \begin{bmatrix} * & * & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \sim \begin{bmatrix} * & * & 0 & 0 \end{bmatrix}$$

and hence

$$\overline{u}_r(0)(\boldsymbol{\gamma}\cdot\mathbf{k})u_s(0)\equiv 0.$$

We also have $k_0 = E$ and

$$\gamma^0 u_s(0) = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix}\begin{bmatrix} * \\ * \\ 0 \\ 0 \end{bmatrix} = u_s(0)$$

so that (using equation (69))

$$\overline{u}_r(k)u_s(k) = 2m\,|c|^2\,(E+m)\overline{u}_r(0)u_s(0) = \overline{u}_r(0)u_s(0)$$

which then implies (choosing the phase equal to +1)

$$c = [2m(E+m)]^{-1/2} \tag{70}$$

In an exactly analogous manner, we have

$$\begin{aligned}\overline{v}_r(k)v_s(k) &= |c'|^2\,\overline{v}_r(0)(\not{k}-m)^2 v_s(0) = |c'|^2\,\overline{v}_r(0)(\not{k}^2 - 2m\not{k} + m^2)v_s(0)\\ &= |c'|^2\,\overline{v}_r(0)(2m^2 - 2mE\gamma^0 - 2m\boldsymbol{\gamma}\cdot\mathbf{k})v_s(0).\end{aligned}$$

But

$$\begin{aligned}\overline{v}_r(0)(\boldsymbol{\gamma}\cdot\mathbf{k})v_s(0) &= \begin{bmatrix} 0 & 0 & * & * \end{bmatrix}\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}\begin{bmatrix} 0 & \boldsymbol{\sigma}\cdot\mathbf{k} \\ -\boldsymbol{\sigma}\cdot\mathbf{k} & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ * \\ * \end{bmatrix}\\ &= \begin{bmatrix} 0 & 0 & * & * \end{bmatrix}\begin{bmatrix} * \\ * \\ 0 \\ 0 \end{bmatrix} \equiv 0\end{aligned}$$

and

$$\overline{v}_r(0)\gamma^0 v_s(0) = \overline{v}_r(0)\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}\begin{bmatrix} 0 \\ * \end{bmatrix} = -\overline{v}_r(0)v_s(0).$$

Therefore

$$\overline{v}_r(k)v_s(k) = |c'|^2\,2m(m+E)\overline{v}_r(0)v_s(0) = \overline{v}_r(0)v_s(0)$$

so that

$$c' = [2m(E+m)]^{-1/2} = c.$$

Lastly, observe that equations (69) and the forms (65) require that we multiply each of equations (65) by $\sqrt{2m}$. We are then left with our final result

$$u_r(k) = \frac{1}{\sqrt{E+m}}(\not{k}+m)\begin{bmatrix}\varphi_r\\0\end{bmatrix} \tag{71a}$$

$$v_r(k) = \frac{-1}{\sqrt{E+m}}(\not{k}-m)\begin{bmatrix}0\\\chi_r\end{bmatrix} \tag{71b}$$

where

$$\varphi_1 = \chi_1 = \begin{bmatrix}1\\0\end{bmatrix} \qquad \text{and} \qquad \varphi_2 = \chi_2 = \begin{bmatrix}0\\1\end{bmatrix}.$$

Explicitly, these may be written out using

$$\not{k}+m = \gamma^0 E - \boldsymbol{\gamma}\cdot\mathbf{k} + mI = \begin{bmatrix}E+m & -\boldsymbol{\sigma}\cdot\mathbf{k}\\ +\boldsymbol{\sigma}\cdot\mathbf{k} & -E+m\end{bmatrix}$$

where

$$\boldsymbol{\sigma}\cdot\mathbf{k} = \begin{bmatrix}k_3 & k_-\\ k_+ & -k_3\end{bmatrix}$$

and

$$\not{k}-m = \gamma^0 E - \boldsymbol{\gamma}\cdot\mathbf{k} - mI = \begin{bmatrix}E-m & -\boldsymbol{\sigma}\cdot\mathbf{k}\\ +\boldsymbol{\sigma}\cdot\mathbf{k} & -(E+m)\end{bmatrix}$$

so that

$$u_r(k) = \frac{1}{\sqrt{E+m}}\begin{bmatrix}E+m & -\boldsymbol{\sigma}\cdot\mathbf{k}\\ \boldsymbol{\sigma}\cdot\mathbf{k} & -E+m\end{bmatrix}\begin{bmatrix}\varphi_r\\0\end{bmatrix} = \frac{1}{\sqrt{E+m}}\begin{bmatrix}(E+m)\varphi_r\\ (\boldsymbol{\sigma}\cdot\mathbf{k})\varphi_r\end{bmatrix}$$

$$v_r(k) = \frac{-1}{\sqrt{E+m}}\begin{bmatrix}E-m & -\boldsymbol{\sigma}\cdot\mathbf{k}\\ \boldsymbol{\sigma}\cdot\mathbf{k} & -(E+m)\end{bmatrix}\begin{bmatrix}0\\\chi_r\end{bmatrix} = \frac{1}{\sqrt{E+m}}\begin{bmatrix}(\boldsymbol{\sigma}\cdot\mathbf{k})\chi_r\\ (E+m)\chi_r\end{bmatrix}$$

or

$$u_1(k) = \frac{1}{\sqrt{E+m}}\begin{bmatrix}E+m\\0\\k_3\\k_+\end{bmatrix} \qquad u_2(k) = \frac{1}{\sqrt{E+m}}\begin{bmatrix}0\\E+m\\k_-\\-k_3\end{bmatrix} \tag{72a}$$

$$v_1(k) = \frac{1}{\sqrt{E+m}}\begin{bmatrix}k_3\\k_+\\E+m\\0\end{bmatrix} \qquad v_2(k) = \frac{1}{\sqrt{E+m}}\begin{bmatrix}k_-\\-k_3\\0\\E+m\end{bmatrix} \tag{72b}$$

Note that to within the normalization constant $1/\sqrt{2m}$, these agree with equations (57) as they should.

Part II – Useful Facts Dealing With the Dirac Spinors

6 Energy Projection Operators and Spin Sums

In order to actually calculate scattering cross sections, there are a number of properties of the Dirac spinors that will prove very useful. We will work in the normalization of Section 5:

$$u_r(k) = \frac{1}{\sqrt{E+m}}(\not{k}+m)\begin{bmatrix}\varphi_r \\ 0\end{bmatrix} \tag{73a}$$

$$v_r(k) = \frac{-1}{\sqrt{E+m}}(\not{k}-m)\begin{bmatrix}0 \\ \chi_r\end{bmatrix} \tag{73b}$$

(these are just equations (71)) where

$$\varphi_1 = \chi_1 = \begin{bmatrix}1 \\ 0\end{bmatrix} \qquad \text{and} \qquad \varphi_2 = \chi_2 = \begin{bmatrix}0 \\ 1\end{bmatrix}$$

and

$$\overline{u}_r(k)u_s(k) = -\overline{v}_r(k)v_s(k) = 2m\delta_{rs} \tag{74}$$

(these are equations (69)). We also have the basic equations (63) and (64)

$$(\not{k}-m)u(\mathbf{k}) = 0 = (\not{k}+m)v(\mathbf{k}) \tag{75a}$$

$$\overline{u}(\mathbf{k})(\not{k}-m) = 0 = \overline{v}(\mathbf{k})(\not{k}+m) \tag{75b}$$

As a consequence of these we immediately have the identity

$$\overline{u}(\mathbf{k})\{(\not{k}-m),\gamma^\mu\}u_s(\mathbf{k}) = 0.$$

Noting that $\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}$ implies $\{\not{k},\gamma^\mu\} = 2k^\mu$, this last equation can be written

$$\overline{u}_r(\mathbf{k})(2k^\mu - 2m\gamma^\mu)u_s(\mathbf{k}) = 2k^\mu\overline{u}_r(\mathbf{k})u_s(\mathbf{k}) - 2m\overline{u}_r(\mathbf{k})\gamma^\mu u_s(\mathbf{k}) = 0.$$

Letting $\mu = 0$ and using (74) yields (where $k^0 = E_{\mathbf{k}} = \omega_{\mathbf{k}}$)

$$u_r^\dagger(\mathbf{k})u_s(\mathbf{k}) = 2\omega_{\mathbf{k}}\delta_{rs}. \tag{76a}$$

Similarly,

$$\overline{v}_r(\mathbf{k})\{(\not{k}+m),\gamma^\mu\}v_s(\mathbf{k}) = \overline{v}_r(\mathbf{k})(2k^\mu + 2m\gamma^\mu)v_s(\mathbf{k}) = 0$$

results in

$$v_r^\dagger(\mathbf{k})v_s(\mathbf{k}) = 2\omega_{\mathbf{k}}\delta_{rs}. \tag{76b}$$

From equation (10) we see that $\not{k}^\dagger = \gamma^{\mu\dagger}k_\mu = \gamma^0\gamma^\mu\gamma^0 k_\mu = \gamma^0\not{k}\gamma^0$ so that from equations (73) we have

$$\begin{aligned}\overline{u}_r(\mathbf{k})v_s(\mathbf{k}) &= \frac{-1}{E+m}\begin{bmatrix}\varphi_r^\dagger & 0\end{bmatrix}(\gamma^0\not{k}\gamma^0+m)\gamma^0(\not{k}-m)\begin{bmatrix}0\\ \chi_s\end{bmatrix}\\ &= \frac{-1}{E+m}\begin{bmatrix}\varphi_r^\dagger & 0\end{bmatrix}\gamma^0(\not{k}+m)(\not{k}-m)\begin{bmatrix}0\\ \chi_s\end{bmatrix}.\end{aligned}$$

There is also a similar result for $\overline{v}_r(\mathbf{k})u_s(\mathbf{k})$, so that using $(\not{k}+m)(\not{k}-m) = k^2-m^2 = 0$ (equation (66)) we obtain

$$\overline{u}_r(\mathbf{k})v_s(\mathbf{k}) = \overline{v}_r(\mathbf{k})u_s(\mathbf{k}) = 0. \tag{77}$$

For convenience, let us define $\widetilde{k}^\mu = (k^0, -\mathbf{k})$ where we still have $k_0^2 - \mathbf{k}^2 = m^2$ or $\omega_{\mathbf{k}} = \omega_{-\mathbf{k}} = \omega_{\widetilde{\mathbf{k}}} = \sqrt{\mathbf{k}^2+m^2}$. From equation (75a) we have $(\widetilde{\not{k}}+m)v_s(-\mathbf{k}) = 0$, so multiplying this from the left by $\overline{u}_r(\mathbf{k})$ and also multiplying $\overline{u}_r(\mathbf{k})(\not{k}-m) = 0$ from the right by $v_s(-\mathbf{k})$ and then adding, we obtain

$$\begin{aligned}0 &= \overline{u}_r(\mathbf{k})(\widetilde{\not{k}}+\not{k})v_s(-\mathbf{k}) = \overline{u}_r(\mathbf{k})(\gamma^0k_0 - \gamma^ik_i + \gamma^0k_0 + \gamma^ik_i)v_s(-\mathbf{k})\\ &= \overline{u}_r(\mathbf{k})2\gamma^0k_0v_s(-\mathbf{k}) = 2\omega_{\mathbf{k}}u_r^\dagger(\mathbf{k})v_s(-\mathbf{k})\end{aligned}$$

so that

$$u_r^\dagger(\mathbf{k})v_s(-\mathbf{k}) = 0 = v_r^\dagger(\mathbf{k})u_s(-\mathbf{k}) \tag{78}$$

where we didn't bother to repeat the identical argument for $v^\dagger u$.

Now for the energy projection operators. First note that because of equations (75) we have $\not{k}u = mu$ and $\not{k}v = -mv$. Then if we define

$$\Lambda^\pm(\mathbf{k}) = \pm\not{k} + m \tag{79}$$

we have

$$\Lambda^+u = \not{k}u + mu = 2mu \quad \text{while} \quad \Lambda^+v = \not{k}v + mv = 0 \tag{80a}$$

and

$$\Lambda^-u = -\not{k}u + mu = 0 \quad \text{while} \quad \Lambda^-v = -\not{k}v + mv = 2mv. \tag{80b}$$

Similarly we find

$$\overline{u}\Lambda^+ = 2m\overline{u} \quad \text{while} \quad \overline{v}\Lambda^+ = 0 \tag{80c}$$

and

$$\overline{u}\Lambda^- = 0 \quad \text{while} \quad \overline{v}\Lambda^- = 2m\overline{v}. \tag{80d}$$

It is easy to see that the $\Lambda^\pm$ are indeed projection operators since

$$(\Lambda^\pm)^2 = (\pm \not{k} + m)^2 = \not{k}^2 \pm 2m\not{k} + m^2 = 2m(\pm\not{k} + m) = 2m\Lambda^\pm \tag{81a}$$

$$\Lambda^\pm \Lambda^\mp = (\pm\not{k} + m)(\mp\not{k} + m) = -\not{k}^2 + m^2 = 0 \tag{81b}$$

$$\Lambda^+ + \Lambda^- = 2mI. \tag{81c}$$

(If we had defined $\Lambda^\pm = (\pm\not{k} + m)/2m$, then these would be $(\Lambda^\pm)^2 = \Lambda^\pm$ and $\Lambda^+ + \Lambda^- = 1$ which is the more common way to define projection operators.)

To put $\Lambda^\pm$ into a more useable form, we proceed as follows. First note that we have $u_r(\mathbf{k}) = S(\Lambda)u_r(\mathbf{0})$ and $v_r(\mathbf{k}) = S(\Lambda)v_r(\mathbf{0})$. Using $S^\dagger\gamma^0 = \gamma^0 S^{-1}$ (equation (40)) these yield $\overline{u}_r(\mathbf{k}) = u_r^\dagger(\mathbf{0})S^\dagger\gamma^0 = \overline{u}_r(\mathbf{0})S^{-1}$ and $\overline{v}_r(\mathbf{k}) = \overline{v}_r(\mathbf{0})S^{-1}$. It will be convenient for us to change notation slightly and write $u(\mathbf{k}, r) = u_r(\mathbf{k})$ so that $u_\alpha(\mathbf{k}, r)$ represents the αth spinor component of $u(\mathbf{k}, r)$. Then consider the sum

$$\begin{aligned}\sum_{r=1}^{2} &[u_\alpha(\mathbf{k}, r)\overline{u}_\beta(\mathbf{k}, r) - v_\alpha(\mathbf{k}, r)\overline{v}_\beta(\mathbf{k}, r)] \\ &= S_{\alpha\mu}\left\{\sum_{r=1}^{2}[u_\mu(\mathbf{0}, r)\overline{u}_\nu(\mathbf{0}, r) - v_\mu(\mathbf{0}, r)\overline{v}_\nu(\mathbf{0}, r)]\right\} S^{-1}_{\nu\beta}. \end{aligned} \tag{82}$$

Observe that what we might call the **outer product** of a column vector $\mathbf{a}$ with a row vector $\mathbf{b}^T$ is

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 & \cdots & a_1 b_n \\ \vdots & & \vdots \\ a_n b_1 & \cdots & a_n b_n \end{bmatrix}$$

so that $(\mathbf{a}\mathbf{b}^T)_{ij} = a_i b_j$. (This is just the usual rule for multiplying matrices. It is also just the usual direct product of two matrices.) To evaluate the term in braces on the right-hand side of equation (82), we first note that from equations (73) we have

$$u(\mathbf{0}, r) = \sqrt{2m}\begin{bmatrix} \varphi_r \\ 0 \end{bmatrix} \quad \text{and} \quad v(\mathbf{0}, r) = \sqrt{2m}\begin{bmatrix} 0 \\ \chi_r \end{bmatrix}$$

and hence the sum of outer products is given by (note γ^0 changes the sign of χ in $\overline{v}$)

$$\begin{aligned} u(\mathbf{0},1)&\overline{u}(\mathbf{0},1) + u(\mathbf{0},2)\overline{u}(\mathbf{0},2) - v(\mathbf{0},1)\overline{v}(\mathbf{0},1) - v(\mathbf{0},2)\overline{v}(\mathbf{0},2) \\ &= 2m\left\{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right. \end{aligned}$$

$$
+\begin{bmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&1&0\\ 0&0&0&0 \end{bmatrix} + \begin{bmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&1 \end{bmatrix} \Bigg\}
$$

$$
= 2mI.
$$

Then the term in braces in equation (82) is just $2m\,\delta_{\mu\nu}$ and we have

$$
\sum_{r=1}^{2}[u_\alpha(\mathbf{k},r)\overline{u}_\beta(\mathbf{k},r) - v_\alpha(\mathbf{k},r)\overline{v}_\beta(\mathbf{k},r)] = 2m\,\delta_{\alpha\beta}.
$$

Taking the (α,β)th element of equation (81c) this last equation can be written as

$$
\sum_{r=1}^{2}[u_\alpha(\mathbf{k},r)\overline{u}_\beta(\mathbf{k},r) - v_\alpha(\mathbf{k},r)\overline{v}_\beta(\mathbf{k},r)] = \Lambda^+_{\alpha\beta} + \Lambda^-_{\alpha\beta}.
$$

Multiplying from the right by $\Lambda^+_{\beta\mu}$ and using equations (80) and (81) we have

$$
\sum_{r=1}^{2}[u_\alpha(\mathbf{k},r)2m\,\overline{u}_\mu(\mathbf{k},r) - 0] = 2m\Lambda^+_{\alpha\mu} + 0
$$

or our desired result

$$
\sum_{r=1}^{2} u_\alpha(\mathbf{k},r)\,\overline{u}_\beta(\mathbf{k},r) = \Lambda^+_{\alpha\beta} = (\not{k} + m)_{\alpha\beta}. \tag{83a}
$$

Similarly, multiplying by $\Lambda^-_{\beta\mu}$ yields

$$
-\sum_{r=1}^{2} v_\alpha(\mathbf{k},r)\,\overline{v}_\beta(\mathbf{k},r) = \Lambda^-_{\alpha\beta} = (-\not{k} + m)_{\alpha\beta}. \tag{83b}
$$

Another way to see this is to use the explicit expressions from the end of Section 5:

$$
\not{k} + m = \begin{bmatrix} E+m & -\boldsymbol{\sigma}\cdot\mathbf{k} \\ +\boldsymbol{\sigma}\cdot\mathbf{k} & -E+m \end{bmatrix} \quad \text{and} \quad \not{k} - m = \begin{bmatrix} E-m & -\boldsymbol{\sigma}\cdot\mathbf{k} \\ +\boldsymbol{\sigma}\cdot\mathbf{k} & -(E+m) \end{bmatrix}
$$

together with

$$
u_r(k) = \sqrt{E+m}\begin{bmatrix} \varphi_r \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\varphi_r \end{bmatrix} \quad \text{and} \quad v_r(k) = \sqrt{E+m}\begin{bmatrix} \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\chi_r \\ \chi_r \end{bmatrix}
$$

so we also have

$$
\begin{aligned}
\overline{u}_r(\mathbf{k}) &= u_r^\dagger(\mathbf{k})\gamma^0 = \sqrt{E+m}\left[\, \varphi_r^\dagger \quad -\varphi_r^\dagger \tfrac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \,\right] \\
\overline{v}_r(\mathbf{k}) &= v_r^\dagger(\mathbf{k})\gamma^0 = \sqrt{E+m}\left[\, \chi_r^\dagger \tfrac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \quad -\chi_r^\dagger \,\right].
\end{aligned}
$$

We now use these to form the outer products

$$u_r(\mathbf{k})\overline{u}_r(\mathbf{k}) = (E+m)\begin{bmatrix} \varphi_r\varphi_r^\dagger & -\varphi_r\varphi_r^\dagger\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\varphi_r\varphi_r^\dagger & -\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\varphi_r\varphi_r^\dagger\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \end{bmatrix}$$

$$v_r(\mathbf{k})\overline{v}_r(\mathbf{k}) = (E+m)\begin{bmatrix} \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\chi_r\chi_r^\dagger\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} & -\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m}\chi_r\chi_r^\dagger \\ \chi_r\chi_r^\dagger\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} & -\chi_r\chi_r^\dagger \end{bmatrix}$$

Noting that

$$\chi_1\chi_1^\dagger + \chi_2\chi_2^\dagger = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

with a similar result for $\varphi_1\varphi_1^\dagger + \varphi_2\varphi_2^\dagger$, we have

$$\sum_{r=1}^{2} u_r(\mathbf{k})\overline{u}_r(\mathbf{k}) = (E+m)\begin{bmatrix} 1 & -\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} & -\frac{(\boldsymbol{\sigma}\cdot\mathbf{k})^2}{(E+m)^2} \end{bmatrix}$$

$$\sum_{r=1}^{2} v_r(\mathbf{k})\overline{v}_r(\mathbf{k}) = (E+m)\begin{bmatrix} \frac{(\boldsymbol{\sigma}\cdot\mathbf{k})^2}{(E+m)^2} & -\frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{k}}{E+m} & -1 \end{bmatrix}.$$

But $(\boldsymbol{\sigma}\cdot\mathbf{k})^2 = \mathbf{k}^2 = E^2 - m^2 = (E+m)(E-m)$ and we are left with

$$\sum_{r=1}^{2} u_r(\mathbf{k})\overline{u}_r(\mathbf{k}) = \begin{bmatrix} E+m & -\boldsymbol{\sigma}\cdot\mathbf{k} \\ \boldsymbol{\sigma}\cdot\mathbf{k} & -E+m \end{bmatrix} = \not{k} + m = \Lambda^+(\mathbf{k})$$

$$\sum_{r=1}^{2} v_r(\mathbf{k})\overline{v}_r(\mathbf{k}) = \begin{bmatrix} E-m & -\boldsymbol{\sigma}\cdot\mathbf{k} \\ \boldsymbol{\sigma}\cdot\mathbf{k} & -(E+m) \end{bmatrix} = \not{k} - m = \Lambda^-(\mathbf{k})$$

which agree with equations (83).

7 Trace Theorems

We now turn to the various trace theorems and some related algebra. Everything is based on the relation

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \tag{84}$$

where

$$g_{\mu\nu} = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix}.$$

We also use the matrix γ_5 defined by (see equation (42))

$$\gamma_5 = \gamma^5 := i\gamma^0\gamma^1\gamma^2\gamma^3 = \frac{i}{4!}\varepsilon_{\alpha\beta\mu\nu}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu. \tag{85}$$

As we saw before, that the last expression is equivalent to $i\gamma^0\gamma^1\gamma^2\gamma^3$ is an immediate consequence of the fact that both $\varepsilon_{\alpha\beta\mu\nu}$ and $\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu$ are antisymmetric in all indices which must also be distinct. Then there are 4! possible orders of indices, and putting them into increasing order introduces the same sign from both terms. This is just a special case of the general result that for any two *antisymmetric* tensors $A_{i_1\cdots i_r}$ and $T^{j_1\cdots j_r}$ we have

$$A_{i_1\cdots i_r}T^{i_1\cdots i_r} = r!A_{|i_1\cdots i_r|}T^{i_1\cdots i_r}$$

where $|i_1\cdots i_r|$ means to sum over all sets of indices with $i_1 < i_2 < \cdots < i_r$.

Now note that

$$\{\gamma_5, \gamma^\mu\} = 0 \tag{86}$$

because

$$\gamma_5\gamma^\mu = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^\mu = -i\gamma^\mu\gamma^0\gamma^1\gamma^2\gamma^3 = -\gamma^\mu\gamma_5$$

due to the fact that no matter what μ is, γ^μ will anticommute with precisely three of the γ's in γ_5, and hence gives an overall $(-)$ sign. We also have

$$(\gamma_5)^2 = 1. \tag{87}$$

To see this, first recall that equation (84) implies $(\gamma^0)^2 = +1$ and $(\gamma^i)^2 = -1$. Then

$$\begin{aligned}(\gamma_5)^2 &= -\gamma^0\gamma^1\gamma^2\gamma^3\gamma^0\gamma^1\gamma^2\gamma^3 = +\gamma^0\gamma^0\gamma^1\gamma^2\gamma^3\gamma^1\gamma^2\gamma^3 = \gamma^1\gamma^2\gamma^3\gamma^1\gamma^2\gamma^3\\ &= \gamma^1\gamma^1\gamma^2\gamma^3\gamma^2\gamma^3 = -\gamma^2\gamma^3\gamma^2\gamma^3 = +\gamma^2\gamma^2\gamma^3\gamma^3 = (-1)^2 = 1.\end{aligned}$$

In what follows, we shall leave off the identity matrix so as to not clutter the equations. For example, to say $\gamma_\mu\gamma^\mu = 4$ really means $\gamma_\mu\gamma^\mu = 4I$.

Theorem 3. *The gamma matrices have the following properties:*

$$\gamma_\mu\gamma^\mu = 4 \tag{88a}$$

$$\gamma_\mu\not{a}\gamma^\mu = -2\not{a} \tag{88b}$$

$$\gamma_\mu\not{a}\not{b}\gamma^\mu = 4a\cdot b \tag{88c}$$

$$\gamma_\mu\not{a}\not{b}\not{c}\gamma^\mu = -2\not{c}\not{b}\not{a} \tag{88d}$$

$$\gamma_\mu\not{a}\not{b}\not{c}\not{d}\gamma^\mu = 2(\not{d}\not{a}\not{b}\not{c} + \not{c}\not{b}\not{a}\not{d}) \tag{88e}$$

$$\{\not{a}, \not{b}\} = 2a\cdot b \tag{88f}$$

Proof. Since $(\gamma^0)^2 = 1$ and $(\gamma^i)^2 = -1$, we have

$$\gamma_\mu \gamma^\mu = (\gamma^0)^2 + \gamma_i \gamma^i = (\gamma^0)^2 - \sum (\gamma^i)^2 = 1 - 3(-1) = 4.$$

Next we have (using (88a))

$$\gamma_\mu \not{a} \gamma^\mu = a_\nu \gamma_\mu \gamma^\nu \gamma^\mu = a_\nu(-\gamma^\nu \gamma_\mu + 2g^\nu_\mu)\gamma^\mu = -\not{a}\gamma_\mu\gamma^\mu + 2\not{a} = -2\not{a}.$$

Note that (88f) simply follows by multiplying $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ by $a_\mu b_\nu$. Then

$$\begin{aligned}
\gamma_\mu \not{a}\not{b}\gamma^\mu &= \gamma_\mu \not{a} b_\nu(-\gamma^\mu\gamma^\nu + 2g^{\mu\nu}) = -\gamma_\mu \not{a}\gamma^\mu \not{b} + 2\not{b}\not{a} \\
&= +2\not{a}\not{b} + 2\not{b}\not{a} && \text{by (88b)} \\
&= 4a \cdot b && \text{by (88f).}
\end{aligned}$$

Now observe that

$$\{\gamma_\mu, \not{a}\} = a^\nu \{\gamma_\mu, \gamma_\nu\} = a^\nu 2g_{\mu\nu} = 2a_\mu.$$

Then

$$\begin{aligned}
\gamma_\mu \not{a}\not{b}\not{c}\gamma^\mu &= \gamma_\mu \not{a}\not{b}(-\gamma^\mu \not{c} + 2c^\mu) = -\gamma_\mu \not{a}\not{b}\gamma^\mu \not{c} + 2\not{c}\not{a}\not{b} \\
&= -4a \cdot b\not{c} + 2\not{c}\not{a}\not{b} && \text{by (88c)} \\
&= -4a \cdot b\not{c} + 2\not{c}(-\not{b}\not{a} + 2a \cdot b) && \text{by (88f)} \\
&= -2\not{c}\not{b}\not{a}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\gamma_\mu \not{a}\not{b}\not{c}\not{d}\gamma^\mu &= \gamma_\mu \not{a}\not{b}\not{c}(-\gamma^\mu \not{d} + 2d^\mu) = -\gamma_\mu \not{a}\not{b}\not{c}\gamma^\mu \not{d} + 2\not{d}\not{a}\not{b}\not{c} \\
&= 2(\not{c}\not{b}\not{a}\not{d} + \not{d}\not{a}\not{b}\not{c}) && \text{by (88d).}
\end{aligned}$$

∎

Theorem 4. *The gamma matrices obey the following trace relations:*

$$\operatorname{tr} I = 4 \tag{89a}$$

$$\operatorname{tr} \gamma_5 = 0 \tag{89b}$$

$$\operatorname{tr}(\text{odd no. } \gamma\text{'s}) = 0 \tag{89c}$$

$$\operatorname{tr} \not{a}\not{b} = 4a \cdot b \tag{89d}$$

$$\operatorname{tr} \not{a}\not{b}\not{c}\not{d} = 4[(a \cdot b)(c \cdot d) + (a \cdot d)(b \cdot c) - (a \cdot c)(b \cdot d)] \tag{89e}$$

$$\operatorname{tr} \gamma_5 \not{a} = 0 \tag{89f}$$

$$\operatorname{tr}\gamma_5 \not{a}\not{b} = 0 \tag{89g}$$

$$\operatorname{tr}\gamma_5 \not{a}\not{b}\not{c} = 0 \tag{89h}$$

$$\operatorname{tr}\gamma_5 \not{a}\not{b}\not{c}\not{d} = -4i\varepsilon^{\mu\nu\rho\sigma} a_\mu b_\nu c_\rho d_\sigma = +4i\varepsilon_{\mu\nu\rho\sigma} a^\mu b^\nu c^\rho d^\sigma \tag{89i}$$

$$\operatorname{tr}\not{a}_1 \cdots \not{a}_{2n} = \operatorname{tr}\not{a}_{2n} \cdots \not{a}_1 \tag{89j}$$

Proof. That $\operatorname{tr} I = 4$ is obvious. Next,

$$\begin{aligned} -i\operatorname{tr}\gamma_5 = \operatorname{tr}\gamma^0\gamma^1\gamma^2\gamma^3 &= -\operatorname{tr}\gamma^1\gamma^2\gamma^3\gamma^0 \quad \text{by (84)} \\ &= +\operatorname{tr}\gamma^1\gamma^2\gamma^3\gamma^0 \quad \text{by the cyclic property of the trace} \end{aligned}$$

Therefore $\operatorname{tr}\gamma_5 = 0$. Now let n be odd and use the fact that $(\gamma_5)^2 = 1$ to write

$$\begin{aligned} \operatorname{tr}\gamma^1 \cdots \gamma^n &= \operatorname{tr}\gamma^1 \cdots \gamma^n \gamma_5\gamma_5 \\ &= -\operatorname{tr}\gamma_5\gamma^1 \cdots \gamma^n\gamma_5 \qquad \text{by (86) with } n \text{ odd} \\ &= -\operatorname{tr}\gamma^1 \cdots \gamma^n(\gamma_5)^2 \qquad \text{by the cyclic property of the trace} \\ &= -\operatorname{tr}\gamma^1 \cdots \gamma^n \end{aligned}$$

so that $\operatorname{tr}\gamma^1 \cdots \gamma^n = 0$. Next we have

$$\begin{aligned} \operatorname{tr}\not{a}\not{b} &= \frac{1}{2}(\operatorname{tr}\not{a}\not{b} + \operatorname{tr}\not{b}\not{a}) = \frac{1}{2}\operatorname{tr}(\not{a}\not{b} + \not{b}\not{a}) \\ &= \frac{1}{2}(2a \cdot b)\operatorname{tr} I = 4a \cdot b \qquad \text{by (88f) and (89a).} \end{aligned}$$

For the next identity we compute

$$\begin{aligned} \operatorname{tr}\not{a}\not{b}\not{c}\not{d} &= -\operatorname{tr}\not{b}\not{a}\not{c}\not{d} + 2a \cdot b \operatorname{tr}\not{c}\not{d} \\ &= +\operatorname{tr}\not{b}\not{c}\not{a}\not{d} - 2a \cdot c \operatorname{tr}\not{b}\not{d} + 8(a \cdot b)(c \cdot d) \\ &= -\operatorname{tr}\not{b}\not{c}\not{d}\not{a} + 2a \cdot d \operatorname{tr}\not{b}\not{c} - 8(a \cdot c)(b \cdot d) + 8(a \cdot b)(c \cdot d) \\ &= -\operatorname{tr}\not{a}\not{b}\not{c}\not{d} + 8(a \cdot d)(b \cdot c) - 8(a \cdot c)(b \cdot d) + 8(a \cdot b)(c \cdot d) \end{aligned}$$

where the first line follows from (88f), the second and third from (88f) and (89d), and the last by the cyclic property of the trace. Hence

$$\operatorname{tr}\not{a}\not{b}\not{c}\not{d} = 4[(a \cdot d)(b \cdot c) - (a \cdot c)(b \cdot d) + (a \cdot b)(c \cdot d)].$$

Equation (89f) follows from (89c) or from $\operatorname{tr}\gamma_5\not{a} = -\operatorname{tr}\not{a}\gamma_5 = -\operatorname{tr}\gamma_5\not{a} = 0$ where we used (86) and the cyclic property of the trace.

For the next result, note that $\operatorname{tr}\gamma_5\not{a}\not{b} = a_\mu b_\nu \operatorname{tr}\gamma_5\gamma^\mu\gamma^\nu$. If $\mu = \nu$ then $\operatorname{tr}\gamma_5(\gamma^\mu)^2 = \pm\operatorname{tr}\gamma_5 = 0$ by (89b). Now assume that $\mu \neq \nu$. Then

$$\gamma_5\gamma^\mu\gamma^\nu = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^\mu\gamma^\nu = \pm i\gamma^\alpha\gamma^\beta$$

where $\alpha \neq \beta$ are those two indices (out of $0, 1, 2, 3$) remaining after moving $\gamma^\mu\gamma^\nu$ through to obtain $(\gamma^\mu)^2$ and $(\gamma^\nu)^2$ (each equal to ± 1). But

$$\operatorname{tr}\gamma^\alpha\gamma^\beta = -\operatorname{tr}\gamma^\beta\gamma^\alpha = -\operatorname{tr}\gamma^\alpha\gamma^\beta = 0$$

if $\alpha \neq \beta$ (by (84) and the cyclic property of the trace). Therefore $\operatorname{tr}\gamma_5\gamma^\mu\gamma^\nu = 0$.

Continuing, we see that (89h) follows from (89c). Now consider $\operatorname{tr}\gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma$. If any two indices are the same, then this is zero since we just showed that $\operatorname{tr}\gamma_5\gamma^\mu\gamma^\nu = 0$. Thus assume that $\mu \neq \nu \neq \rho \neq \sigma$, i.e., (μ, ν, ρ, σ) is a permutation of $(0, 1, 2, 3)$. If we choose $(\mu, \nu, \rho, \sigma) = (0, 1, 2, 3)$ then we have

$$\operatorname{tr}\gamma_5\gamma^0\gamma^1\gamma^2\gamma^3 = -i\operatorname{tr}(\gamma_5)^2 = -4i = -4i\varepsilon^{0123}.$$

Since both sides of this equation are totally antisymmetric, it must hold for all values of the indices so that

$$\operatorname{tr}\gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = -4i\varepsilon^{\mu\nu\rho\sigma}$$

and hence

$$\operatorname{tr}\gamma_5 \not{a}\not{b}\not{c}\not{d} = -4i\varepsilon^{\mu\nu\rho\sigma}a_\mu b_\nu c_\rho d_\sigma.$$

For the second part of (89i), note that $\varepsilon_{\mu\nu\rho\sigma}$ is *not* a tensor by definition. Indeed, we have $\varepsilon_{0123} := \varepsilon^{0123} := +1$. But the only non-vanishing terms in $\varepsilon^{\mu\nu\rho\sigma}a_\mu b_\nu c_\rho d_\sigma$ comes when all the indices are distinct, and $a_0 = a^0$ while $a_i = -a^i$ etc., so we actually have

$$\varepsilon^{\mu\nu\rho\sigma}a_\mu b_\nu c_\rho d_\sigma = -\varepsilon_{\mu\nu\rho\sigma}a^\mu b^\nu c^\rho d^\sigma.$$

Finally, let us define $\widetilde{\gamma}^\mu = -(\gamma^\mu)^T$. Then $\{\widetilde{\gamma}^\mu, \widetilde{\gamma}^\nu\} = 2g^{\mu\nu}$, and hence by Theorem 2 there exists a matrix C such that $C\gamma^\mu C^{-1} = -(\gamma^\mu)^T$. Then we have

$$\begin{aligned}
\operatorname{tr}\not{a}_1\cdots\not{a}_{2n} &= \operatorname{tr}C\not{a}_1\cdots\not{a}_{2n}C^{-1} = \operatorname{tr}C\not{a}C^{-1}C\not{a}_2\cdots C\not{a}_{2n}C^{-1} \\
&= (-1)^{2n}\operatorname{tr}\not{a}_1^T\cdots\not{a}_{2n}^T = \operatorname{tr}(\not{a}_{2n}\cdots\not{a}_1)^T \\
&= \operatorname{tr}\not{a}_{2n}\cdots\not{a}_1.
\end{aligned}$$

∎

Part III – Group Theoretic Approach

8 Decomposing the Lorentz Group

In this part we will derive the Dirac equation from the standpoint of representations of the Lorentz group. To begin with, let us first show that the (homogeneous) Lorentz transformations form a group. (By "homogeneous" we mean as distinct from the general Poincaré (or inhomogeneous Lorentz) transformations of the form $x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu$ where a^ν is a constant.)

Clearly the set of all Lorentz transformations contains the identity transformation (i.e., zero boost), and corresponding to each Lorentz transformation $\Lambda^\mu{}_\nu$ there is an inverse transformation $\Lambda^{-1} = \Lambda^T$ (equation (28)). Thus we need only show that the set of Lorentz transformations is closed, i.e., that the composition of two Lorentz transformations is another such transformation. Well, if $x'^\mu = \Lambda^\mu{}_\alpha x^\alpha$ and $x''^\nu = \Lambda'^\nu{}_\beta x'^\beta$, then

$$x''^\nu = \Lambda'^\nu{}_\beta x'^\beta = \Lambda'^\nu{}_\beta \Lambda^\beta{}_\alpha x^\alpha$$

and therefore (using equation (30))

$$\begin{aligned} x''^\nu x''_\nu &= \Lambda'^\nu{}_\beta \Lambda^\beta{}_\alpha x^\alpha \Lambda'_{\nu\rho} \Lambda^{\rho\sigma} x_\sigma = (\Lambda'^\nu{}_\beta \Lambda'_{\nu\rho})(\Lambda^\beta{}_\alpha \Lambda^{\rho\sigma}) x^\alpha x_\sigma \\ &= g_{\beta\rho}(\Lambda^\beta{}_\alpha \Lambda^{\rho\sigma}) x^\alpha x_\sigma = \Lambda_{\rho\alpha}\Lambda^{\rho\sigma} x^\alpha x_\sigma = g^\sigma_\alpha x^\alpha x_\sigma = x^\sigma x_\sigma \end{aligned}$$

and therefore the composition of two Lorentz transformations is also a Lorentz transformation (because it preserves the length $x^\sigma x_\sigma$). This defines the **Lorentz group**.

If we let x^μ denote the "lab frame" and x'^μ the "moving frame," then for a pure boost along the x^1-axis we have (note that this is just equation (49) with $\beta \to -\beta$ and switching the primes)

$$x^0 = \gamma(x'^0 + \beta x'^1) \qquad x^1 = \gamma(x'^1 + \beta x'^0) \qquad x^2 = x'^2 \qquad x^3 = x'^3$$

with the corresponding Lorentz transformation matrix

$$\Lambda(\beta)^\mu{}_\nu = \begin{bmatrix} \gamma & \gamma\beta & & \\ \gamma\beta & \gamma & & \\ & & 1 & \\ & & & 1 \end{bmatrix}. \tag{90}$$

However, we can also include purely spatial rotations since these also preserve the lengths $\mathbf{x} \cdot \mathbf{x}$ and hence also $x^\mu x_\mu = (x^0)^2 - \mathbf{x} \cdot \mathbf{x}$. For example, a purely spatial rotation about the x^3-axis has the Lorentz transformation matrix

$$\Lambda(\theta)^\mu{}_\nu = \begin{bmatrix} 1 & & & \\ & \cos\theta & -\sin\theta & \\ & \sin\theta & \cos\theta & \\ & & & 1 \end{bmatrix}. \tag{91}$$

Since such rotations clearly form a group themselves, the spatial rotations form a subgroup of the Lorentz group. However, the set of pure boosts *does not* form a subgroup. In fact, the commutator of two different boost generators is a rotation generator (see equation (93b) below), and this is the origin of Thomas precession.

It is easy to see that the Lorentz group depends on six real parameters. From a physical viewpoint, there can be three independent boost directions, so there will be three boost parameters (like the direction cosines in equation (52)), and there are obviously three independent rotation angles for another three parameters (for example, the three Euler angles). Alternatively, Lorentz transformations are also defined by the condition $\Lambda^T g \Lambda = g$. Since both sides of this equation are symmetric 4×4 matrices, there are $(4^2 - 4)/2 + 4 = 10$ constraint equations coupled with the 16 entries in a 4×4 matrix for a total of 6 independent entries.

For an infinitesimal boost $\beta \ll 1$, equation (90) becomes

$$\Lambda(\beta)^\mu{}_\nu = \begin{bmatrix} 1 & \beta & & \\ \beta & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} := I + \omega_{10} M^{10}$$

where $I = (g^\mu{}_\nu) = (\delta^\mu_\nu)$, the infinitesimal boost parameter along the x-axis is defined by $\omega_{10} := \beta$, and we define the generator of boosts along the x-axis by

$$M^{10} = \begin{bmatrix} 0 & 1 & & \\ 1 & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}.$$

Let us *define* $\omega_{01} := -\omega_{10}$ and $M^{01} := -M^{10}$ so that

$$\omega_{10} M^{10} = \frac{1}{2}(\omega_{10} M^{10} + \omega_{10} M^{10}) = \frac{1}{2}(\omega_{10} M^{10} + \omega_{01} M^{01}).$$

Since we can clearly do the same thing for boosts along the x^2 and x^3 directions, we see that the general boost generator matrix (as in equation (52)) is given by

$$\frac{1}{2}(\omega_{i0} M^{i0} + \omega_{0i} M^{0i})$$

where

$$M^{20} = \begin{bmatrix} 0 & & 1 & \\ & 0 & & \\ 1 & & 0 & \\ & & & 0 \end{bmatrix} \qquad \text{and} \qquad M^{30} = \begin{bmatrix} 0 & & & 1 \\ & 0 & & \\ & & 0 & \\ 1 & & & 0 \end{bmatrix}.$$

Similarly, for an infinitesimal spatial rotation $\theta \ll 1$, equation (91) becomes

$$\Lambda(\theta)^{\mu}{}_{\nu} = \begin{bmatrix} 1 & & & \\ & 1 & -\theta & \\ & \theta & 1 & \\ & & & 1 \end{bmatrix} := I + \omega_{12} M^{12}$$

where now the parameter for rotation about the x^3-axis is defined by $\omega_{12} := \theta$, and the corresponding rotation generator is defined by

$$M^{12} = \begin{bmatrix} 0 & & & \\ & 0 & -1 & \\ & 1 & 0 & \\ & & & 0 \end{bmatrix}.$$

Again, we define $\omega_{21} = -\omega_{12}$ and $M^{21} = -M^{12}$ so that

$$\omega_{12} M^{12} = \frac{1}{2}(\omega_{12} M^{12} + \omega_{12} M^{12}) = \frac{1}{2}(\omega_{12} M^{12} + \omega_{21} M^{21}).$$

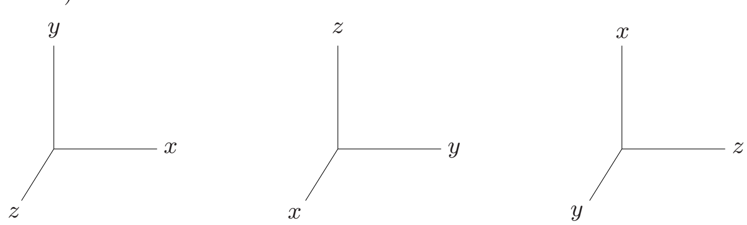
With

$$M^{23} = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & -1 \\ & & 1 & 0 \end{bmatrix} \quad \text{and} \quad M^{31} = \begin{bmatrix} 0 & & & \\ & 0 & & 1 \\ & & 0 & \\ & -1 & & 0 \end{bmatrix}$$

as the generators of rotations about the x^1 and x^2 axes respectively, we have the spatial rotation generator matrix

$$\frac{1}{2}(\omega_{ij} M^{ij} + \omega_{ji} M^{ji}).$$

(Notice that the signs appear wrong in M^{31}, but they're not. This is because all rotations are defined by the right-handed orientation of $\mathbb{R}^3$ as shown in the figure below.)



Just as we did in equation (53), the matrix representing a finite Lorentz transformation (including both boosts and rotations) is obtained by exponentiating the infinitesimal results so that we have the general result (where we define $\omega_{00} = \omega_{ii} = 0$)

$$\Lambda = e^{\frac{1}{2}\omega_{\mu\nu} M^{\mu\nu}}. \tag{92}$$

Let us define the vectors

$$\mathbf{M} = (M^{23}, M^{31}, M^{12}) \qquad \text{and} \qquad \mathbf{N} = (M^{10}, M^{20}, M^{30}).$$

By direct computation you can easily show that (using the summation convention on repeated letters)

$$[M_i, M_j] = \varepsilon_{ijk} M_k \tag{93a}$$

$$[N_i, N_j] = -\varepsilon_{ijk} M_k \tag{93b}$$

$$[M_i, N_j] = \varepsilon_{ijk} N_k. \tag{93c}$$

Note that these are now to be interpreted as the components of vectors in $\mathbb{R}^3$ where the metric is just δ_{ij}. In other words, they are a shorthand for relations such as $[M^{23}, M^{31}] = M^{12}$ plus its cyclic permutations. They are exactly the same as you have seen in elementary treatments of angular momentum in quantum mechanics. It is also worth observing that from equation (93b) we see that a combination of two boosts in different directions results in a rotation rather than another boost. As mentioned above, this is in fact the origin of Thomas precession.

However, we see that the vectors $\mathbf{M}$ and $\mathbf{N}$ don't commute, and hence we define the new vectors

$$\mathbf{J} = \frac{i}{2}(\mathbf{M} + i\mathbf{N}) \qquad \text{and} \qquad \mathbf{K} = \frac{i}{2}(\mathbf{M} - i\mathbf{N})$$

so that

$$\mathbf{M} = -i(\mathbf{J} + \mathbf{K}) \qquad \text{and} \qquad \mathbf{N} = \mathbf{K} - \mathbf{J}. \tag{94}$$

Now you can also easily show that

$$[J_i, J_j] = i\varepsilon_{ijk} J_k \tag{95a}$$

$$[K_i, K_j] = i\varepsilon_{ijk} K_k \tag{95b}$$

$$[J_i, K_j] = 0 \tag{95c}$$

which are simply the commutation relations for two sets of independent, commuting angular momentum generators of the group $SU(2)$. Thus we have shown that a general Lorentz transformation is of the equivalent forms

$$\Lambda = e^{\omega_{i0} M^{i0} + \frac{1}{2}\omega_{ij} M^{ij}} = e^{\mathbf{a}\cdot\mathbf{N} + \mathbf{b}\cdot\mathbf{M}} = e^{-(\mathbf{a}+i\mathbf{b})\cdot\mathbf{J} + (\mathbf{a}-i\mathbf{b})\cdot\mathbf{K}}. \tag{96}$$

And since $\mathbf{J}$ and $\mathbf{K}$ commute, this is

$$\Lambda = e^{-(\mathbf{a}+i\mathbf{b})\cdot\mathbf{J}} e^{(\mathbf{a}-i\mathbf{b})\cdot\mathbf{K}}. \tag{97}$$

Note that since $\mathbf{M}$ and $\mathbf{N}$ are real and antisymmetric, it is easy to see that $\mathbf{J}^\dagger = \mathbf{K}$ so that $\Lambda^\dagger = \Lambda^T = \Lambda^{-1}$ which again shows that Lorentz transformations are orthogonal.

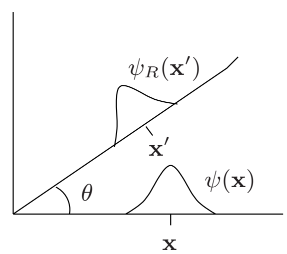
9 Angular Momentum in Quantum Mechanics

While the decomposition (97) may look different, it is actually exactly the same as what you have probably learned in quantum mechanics when you studied the addition of angular momentum. So, to help clarify what we have done, let's briefly review the theory of representations of the rotation group as applied to angular momentum in quantum mechanics.

First, let's take a brief look at how the spatial rotation operator is defined. If we rotate a vector $\mathbf{x}$ in $\mathbb{R}^3$, then we obtain a new vector $\mathbf{x}' = R(\boldsymbol{\theta})\mathbf{x}$ where $R(\boldsymbol{\theta})$ is the matrix that represents the rotation. In two dimensions this is

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

If we have a scalar wavefunction $\psi(\mathbf{x})$, then under rotation we obtain a new wavefunction $\psi_R(\mathbf{x}')$, where $\psi(\mathbf{x}) = \psi_R(\mathbf{x}') = \psi_R(R(\boldsymbol{\theta})\mathbf{x})$. (See the figure below.)



Alternatively, we can write

$$\psi_R(\mathbf{x}) = \psi(R^{-1}(\boldsymbol{\theta})\mathbf{x}).$$

Since R is an orthogonal transformation (it preserves the length of $\mathbf{x}$) we know that $R^{-1}(\boldsymbol{\theta}) = R^T(\boldsymbol{\theta})$, and in the case where $\theta \ll 1$ we then have

$$R^{-1}(\boldsymbol{\theta})\mathbf{x} = \begin{bmatrix} x + \theta y \\ -\theta x + y \end{bmatrix}.$$

Expanding $\psi(R^{-1}(\boldsymbol{\theta})\mathbf{x})$ with these values for x and y we have

$$\psi_R(\mathbf{x}) = \psi(x + \theta y, y - \theta x) = \psi(\mathbf{x}) - \theta[x\partial_y - y\partial_x]\psi(\mathbf{x})$$

or, using $p^i = -i\partial_i$ this is

$$\psi_R(\mathbf{x}) = \psi(\mathbf{x}) - i\theta[xp_y - yp_x]\psi(\mathbf{x}) = [1 - i\theta L_z]\psi(\mathbf{x}).$$

For finite θ we exponentiate this to write $\psi_R(\mathbf{x}) = e^{-i\theta L_z}\psi(\mathbf{x})$, and in the case of an arbitrary angle $\boldsymbol{\theta}$ in $\mathbb{R}^3$ this becomes

$$\psi_R(\mathbf{x}) = e^{-i\boldsymbol{\theta}\cdot\mathbf{L}}\psi(\mathbf{x}).$$

In an abstract notation we write this as

$$|\psi_R\rangle = U(R)|\psi\rangle$$

where $U(R) = e^{-i\boldsymbol{\theta}\cdot\mathbf{L}}$. For simplicity and clarity, we have written $U(R)$ rather than the more complete $U(R(\boldsymbol{\theta}))$ which we continue to do unless the more complete notation is needed.

What we just did was for orbital angular momentum. In the case of spin there is no classical counterpart, so we *define* the spin angular momentum operator $\mathbf{S}$ to obey the usual commutation relations, and the spin states to transform under the rotation operator $e^{-i\boldsymbol{\theta}\cdot\mathbf{S}}$. It is common to use the symbol $\mathbf{J}$ to stand for any type of angular momentum operator, for example $\mathbf{L}, \mathbf{S}$ or $\mathbf{L}+\mathbf{S}$, and this is what we shall do from now on. The operator $\mathbf{J}$ is called the **total angular momentum operator**. (This example applied to a scalar wavefunction ψ, which represents a spinless particle. Particles with spin are described by vector wavefunctions $\boldsymbol{\psi}$ (as we have seen for the Dirac spinors), and in this case the spin operator $\mathbf{S}$ serves to mix up the components of $\boldsymbol{\psi}$ under rotations.)

The angular momentum operators $J^2 = \mathbf{J}\cdot\mathbf{J}$ and J_z commute and hence have the simultaneous eigenstates denoted by $|jm\rangle$ with the property that (with $\hbar = 1$)

$$J^2|jm\rangle = j(j+1)|jm\rangle \qquad \text{and} \qquad J_z|jm\rangle = m|jm\rangle$$

where m takes the $2j+1$ values $-j \le m \le j$. Since the rotation operator is given by $U(R) = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}}$ we see that $[U(R), J^2] = 0$. Then

$$J^2U(R)|jm\rangle = U(R)J^2|jm\rangle = j(j+1)U(R)|jm\rangle$$

so that the *magnitude* of the angular momentum can't change under rotations. However, $[U(R), J_z] \ne 0$ so the rotated state will no longer be an eigenstate of J_z with the same eigenvalue m.

Note that acting to the right we have the matrix element

$$\langle j'm'|J^2U(R)|jm\rangle = \langle j'm'|U(R)J^2|jm\rangle = j(j+1)\langle j'm'|U(R)|jm\rangle$$

while acting to the left gives

$$\langle j'm'|J^2U(R)|jm\rangle = j'(j'+1)\langle j'm'|U(R)|jm\rangle$$

and therefore

$$\langle j'm'|U(R)|jm\rangle = 0 \quad \text{unless } j = j'. \tag{98}$$

We also make note of the fact that acting with J^2 and J_z in both directions yields

$$\langle j'm'|J^2|jm\rangle = j'(j'+1)\langle j'm'|jm\rangle = j(j+1)\langle j'm'|jm\rangle$$

and

$$\langle jm'|J_z|jm\rangle = m'\langle jm'|jm\rangle = m\langle jm'|jm\rangle$$

so that (as you should have already known)

$$\langle j'm'|jm\rangle = \delta_{j'j}\delta_{m'm}. \tag{99}$$

In other words, the states $|jm\rangle$ form a complete orthonormal set, and the state $U(R)|jm\rangle$ must be of the form

$$U(R)|jm\rangle = \sum_{m'} |jm'\rangle\langle jm'|U(R)|jm\rangle = \sum_{m'} |jm'\rangle \mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta}) \tag{100}$$

where

$$\mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta}) := \langle jm'|U(R)|jm\rangle = \langle jm'|e^{-i\boldsymbol{\theta}\cdot\mathbf{J}}|jm\rangle. \tag{101}$$

(Notice the order of susbscripts in the sum in equation (100). This is the same as the usual definition of the matrix representation $[T]_e = (a_{ij})$ of a linear operator $T : V \to V$ defined by $Te_i = \sum_j e_j a_{ji}$.)

Since for each j there are $2j+1$ values of m, we have constructed a $(2j+1) \times (2j+1)$ matrix $\mathscr{D}^{(j)}(\boldsymbol{\theta})$ for each value of j. This matrix is referred to as the jth **irreducible representation** of the rotation group. The word "irreducible" means that there is no subset of the space of states $\{|jj\rangle, |j, m-1\rangle, \ldots, |j, -j\rangle\}$ that transforms into itself under *all* rotations $U(R(\boldsymbol{\theta}))$. Put in another way, a representation is irreducible if the vector space on which it acts has no invariant subspaces.

Now, it is a general result of the theory of group representations that any representation of a finite group or compact Lie group is equivalent to a unitary representation, and any reducible unitary representation is completely reducible. Therefore, any representation of a finite group or compact Lie group is either already irreducible or else it is completely reducible (i.e., the space on which the operators act can be put into block diagonal form where each block corresponds to an invariant subspace). However, at this point we don't want to get into the general theory of representations, so let us prove directly that the representations $\mathscr{D}^{(j)}(\boldsymbol{\theta})$ of the rotation group are irreducible. Recall that the raising and lowering operators $J_\pm$ are defined by

$$J_\pm|jm\rangle = (J_x \pm iJ_y)|jm\rangle = \sqrt{j(j+1) - m(m\pm 1)}\,|j, m\pm 1\rangle.$$

In particular, the operators $J_\pm$ don't change the value of j when acting on the states $|jm\rangle$.

Theorem 5. *The representations $\mathscr{D}^{(j)}(\boldsymbol{\theta})$ of the rotation group are irreducible. In other words, there is no subspace of the space of states $|jm\rangle$ (for fixed j) that transforms among itself under all rotations.*

Proof. Fix j and let V be the space spanned by the $2j+1$ vectors $|jm\rangle := |m\rangle$. We claim that V is irreducible with respect to rotations $U(R)$. This means that given *any* $|u\rangle \in V$, the set of all vectors of the form $U(R)|u\rangle$ (i.e., for *all* rotations $U(R)$) spans V. (Otherwise, if there exists $|v\rangle$ such that $\{U(R)|v\rangle\}$ didn't span V, then V would be reducible since the collection of all such $U(R)|v\rangle$ would define an invariant subspace.)

To show V is irreducible, let $\widetilde{V} = \text{span}\{U(R)|u\rangle\}$ where $|u\rangle \in V$ is arbitrary but fixed. For infinitesimal $\boldsymbol{\theta}$ we have $U(R(\boldsymbol{\theta})) = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}} = 1 - i\boldsymbol{\theta}\cdot\mathbf{J}$ and in particular $U(R(\varepsilon\widehat{\mathbf{x}})) = 1 - i\varepsilon J_x$ and $U(R(\varepsilon\widehat{\mathbf{y}})) = 1 - i\varepsilon J_y$. Then

$$\begin{aligned}
J_\pm|u\rangle = (J_x \pm iJ_y)|u\rangle &= \left\{\frac{1}{i\varepsilon}[1 - U(R(\varepsilon\widehat{\mathbf{x}}))] \pm i\left(\frac{1}{i\varepsilon}[1 - U(R(\varepsilon\widehat{\mathbf{y}}))]\right)\right\}|u\rangle \\
&= \frac{1}{\varepsilon}\big\{\pm[1 - U(R(\varepsilon\widehat{\mathbf{y}}))] - i + iU(R(\varepsilon\widehat{\mathbf{x}}))\big\}|u\rangle \in \widetilde{V}
\end{aligned}$$

by definition of $\widetilde{V}$ and vector spaces. Since $J_\pm$ acting on $|u\rangle$ is a linear combination of rotations acting on $|u\rangle$ and this is in $\widetilde{V}$, we see that $(J_\pm)^2$ acting on $|u\rangle$ is again some other linear combination of rotations acting on $|u\rangle$ and hence is also in $\widetilde{V}$. So in general, we see that $(J_\pm)^n|u\rangle$ is again in $\widetilde{V}$.

By definition of V, we may write (since j is fixed)

$$\begin{aligned}
|u\rangle &= \sum_m |jm\rangle\langle jm|u\rangle = \sum_m |m\rangle\langle m|u\rangle \\
&= |\overline{m}\rangle\langle\overline{m}|u\rangle + |\overline{m}+1\rangle\langle\overline{m}+1|u\rangle + \cdots + |j\rangle\langle j|u\rangle
\end{aligned}$$

where $\overline{m}$ is simply the smallest value of m for which $\langle\overline{m}|u\rangle \neq 0$ (and not all of the terms up to $\langle j|u\rangle$ are necessarily nonzero). Acting on this with J_+ we obtain (leaving off the constant factors and noting that $J_+|j\rangle = 0$)

$$J_+|u\rangle \sim |\overline{m}+1\rangle\langle\overline{m}|u\rangle + |\overline{m}+2\rangle\langle\overline{m}+1|u\rangle + \cdots + |j\rangle\langle j-1|u\rangle \in \widetilde{V}.$$

Since $\langle\overline{m}|u\rangle \neq 0$ by assumption, it follows that $|\overline{m}+1\rangle \in \widetilde{V}$.

We can continue to act on $|u\rangle$ with J_+ a total of $j - \overline{m}$ times at which point we will have shown that $|\overline{m} + j - \overline{m}\rangle = |j\rangle := |jj\rangle \in \widetilde{V}$. Now we can apply J_- $2j+1$ times to $|jj\rangle$ to conclude that the $2j+1$ vectors $|jm\rangle$ all belong to $\widetilde{V}$, and thus $\widetilde{V} = V$. (This is because we have really just applied the combination of rotations $(J_-)^{2j+1}(J_+)^{j-\overline{m}}$ to $|u\rangle$, and each step along the way is just some vector in $\widetilde{V}$.) ∎

Now suppose that we have two angular momenta $\mathbf{J}_1$ and $\mathbf{J}_2$ with $[\mathbf{J}_1, \mathbf{J}_2] = 0$. Then ${J_1}^2, {J_2}^2, J_{1z}$ and J_{2z} all commute, and hence we can construct simultaneous eigenstates which we write as $|j_1 j_1 m_1 m_2\rangle$. Furthermore, with $\mathbf{J} = \mathbf{J_1} + \mathbf{J_2}$ it follows that $J^2, J_z, {J_1}^2$ and ${J_2}^2$ all commute, and hence we can also construct the simultaneous eigenstates $|j_1 j_2 jm\rangle$. However, J^2 does not commute with either J_{1z} or J_{2z}, so if we specify J^2, then we are only free to also specify $J_z = J_{1z} + J_{2z}$ and not J_{1z} or J_{2z} individually.

Since j_1 and j_2 are fixed, there are $2j_1+1$ possible values for m_1, and $2j_2+1$ possible values of m_2 for a total of $(2j_1+1)(2j_2+1)$ linearly independent states of the form $|j_1 j_2 m_1 m_2\rangle$, which must be the same as the number of states of the form $|j_1 j_2 jm\rangle$. The maximum possible value of $m = m_1 + m_2$ is $j_1 + j_2$. But the next value $j_1 + j_2 - 1$ can be due to either $m_1 = j_1$ and $m_2 = j_2 - 1$, or due to $m_1 = j_1 - 1$ and $m_2 = j_2$. And as we go to lower values of m, there will be even more possible choices. Since the largest m value is j_1+j_2, there must be a state of total $j=j_1+j_2$, and a multiplet of $2(j_1+j_2)+1$ states corresponding to this value of j. But this $j=j_1+j_2$ state only accounts for one of the states with $m=j_1+j_2-1$, and hence there must be another state with $j=j_1+j_2-1$ and its corresponding multiplet of $2(j_1+j_2-1)+1$ states $|j_1j_2jm\rangle$.

We can continue to consider states with lower and lower j values and corresponding multiplets of $2j+1$ possible m values. However, the total number of states must equal $(2j_1+1)(2j_2+1)$, and hence we have the relation

$$\sum_{j=N}^{j_1+j_1} 2j+1 = (2j_1+1)(2j_2+1)$$

where N is to be determined. Using the formula (which you can check)

$$\sum_{j=n_1}^{n_2} j = \frac{1}{2}(n_2-n_1+1)(n_2+n_1)$$

it takes a little bit of elementary algebra to show that $N^2=(j_1-j_2)^2$ and therefore the minimum value of j is $|j_1-j_2|$, i.e., $|j_1-j_2|\le j\le j_1+j_2$.

What else can we say about the angular momentum states $|j_1j_2m_1m_2\rangle$? These are really direct product states and are written in the equivalent forms

$$|j_1j_2m_1m_2\rangle = |j_1m_1\rangle\otimes|j_2m_2\rangle = |j_1m_1\rangle|j_2m_2\rangle.$$

(Remark: You may recall from linear algebra that given two vector spaces V and V', we may define a bilinear map $V\times V'\to V\otimes V'$ that takes ordered pairs $(v,v')\in V\times V'$ and gives a new vector denoted by $v\otimes v'$. Since this map is bilinear by definition, if we have the linear combinations $v=\sum x_iv_i$ and $v'=\sum y_jv'_j$ then $v\otimes v'=\sum x_iy_j(v_i\otimes v'_j)$. In particular, if V has basis $\{e_i\}$ and V' has basis $\{e'_j\}$, then $\{e_i\otimes e'_j\}$ is a basis for $V\otimes V'$ which is then of dimension $(\dim V)(\dim V')$ and called the **direct** (or **tensor**) **product** of V and V'. Then, if we are given two operators $A\in L(V)$ and $B\in L(V')$, the **direct product** of A and B is the operator $A\otimes B$ defined on $V\otimes V'$ by $(A\otimes B)(v\otimes v') := A(v)\otimes B(v')$.)

When we write $\langle j_1j_2m_1m_2|j'_1j'_2m'_1m'_2\rangle$ we really mean

$$(\langle j_1m_1|\otimes\langle j_2m_2|)(|j'_1m'_1\rangle\otimes|j'_2m'_2\rangle) = \langle j_1m_1|j'_1m'_1\rangle\langle j_2m_2|j'_2m'_2\rangle$$

so that by equation (99) we have

$$\langle j_1j_2m_1m_2|j'_1j'_2m'_1m'_2\rangle = \delta_{j_1j'_1}\delta_{j_2j'_2}\delta_{m_1m'_1}\delta_{m_2m'_2} \tag{102}$$

We point out that a special case of this result is

$$\langle j_1j_2m_1m_2|j_1j_2m'_1m'_2\rangle = \delta_{m_1m'_1}\delta_{m_2m'_2}$$

As we will see below, the matrix elements $\langle j_1j_2jm|j_1j_2m_1m_2\rangle$ will be very important to us. Now consider the expression $J_z=J_{1z}+J_{2z}$. What the operator on the right really stands for is $J_{1z} \otimes I + I \otimes J_{2z}$. Then on the one hand we have

$$\begin{aligned}
\langle j_1 j_2 jm | J_z | j_1 j_2 m_1 m_2 \rangle &= \langle j_1 j_2 jm | (J_{1z} \otimes I + I \otimes J_{2z})(|j_1 m_1\rangle \otimes |j_2 m_2\rangle) \\
&= \langle j_1 j_2 jm | (J_{1z} |j_1 m_1\rangle \otimes |j_2 m_2\rangle + |j_1 m_1\rangle \otimes J_{2z} |j_2 m_2\rangle) \\
&= (m_1 + m_2) \langle j_1 j_2 jm | (|j_1 m_1\rangle \otimes |j_2 m_2\rangle) \\
&= (m_1 + m_2) \langle j_1 j_2 jm | j_1 j_2 m_1 m_2 \rangle
\end{aligned}$$

while acting to the left with J_z we have

$$\langle j_1 j_2 jm | J_z | j_1 j_2 m_1 m_2 \rangle = m \langle j_1 j_2 jm | j_1 j_2 m_1 m_2 \rangle .$$

Comparing these last two results shows that

$$\langle j_1 j_2 jm | j_1 j_2 m_1 m_2 \rangle = 0 \quad \text{unless } m = m_1 + m_2 . \tag{103}$$

We also see that for $i = 1, 2$ we can let ${J_i}^2$ act either to the right or left also and we have

$$\begin{aligned}
\langle j_1' j_2' jm | {J_i}^2 | j_1 j_2 m_1 m_2 \rangle &= j_i (j_i + 1) \langle j_1' j_2' jm | j_1 j_2 m_1 m_2 \rangle \\
&= j_i' (j_i' + 1) \langle j_1' j_2' jm | j_1 j_2 m_1 m_2 \rangle
\end{aligned}$$

so that

$$\langle j_1' j_2' jm | j_1 j_2 m_1 m_2 \rangle = 0 \quad \text{unless } j_1' = j_1 \text{ and } j_2' = j_2 . \tag{104}$$

In a similar manner, by letting the operator act to both the right and left we can consider the matrix elements of ${J_1}^2, {J_2}^2, J^2$ and J_z to show that (compare with equation (102))

$$\langle j_1' j_2' j' m' | j_1 j_2 jm \rangle = \delta_{j_1' j_1} \delta_{j_2' j_2} \delta_{j' j} \delta_{m' m} \tag{105}$$

Since both of the sets $\{|j_1 j_2 m_1 m_2\rangle\}$ and $\{|j_1 j_2 jm\rangle\}$ are complete (i.e., they form a basis for the two particle angular momentum states where j_1 and j_2 are fixed), we can write either set as a function of the other:

$$|j_1 j_2 jm\rangle = \sum_{\substack{m_1, m_2 \\ (m_1 + m_2 = m)}} |j_1 j_2 m_1 m_2\rangle \langle j_1 j_2 m_1 m_2 | j_1 j_2 jm \rangle \tag{106}$$

or

$$|j_1 j_2 m_1 m_2\rangle = \sum_{\substack{j \\ (m = m_1 + m_2)}} |j_1 j_2 jm\rangle \langle j_1 j_2 jm | j_1 j_2 m_1 m_2 \rangle . \tag{107}$$

The matrix elements $\langle j_1 j_2 jm | j_1 j_2 m_1 m_2 \rangle$ are called **Clebsch-Gordan coefficients**, and may always be taken to be real (see below). Note that the Clebsch-Gordan coefficients are nothing more than the matrix elements of the transition matrix that changes between the $|j_1 j_2 jm\rangle$ basis and the $|j_1 j_2 m_1 m_2\rangle$ basis. In fact, since both bases are orthonormal, the Clebsch-Gordan coefficients define a unitary transformation matrix.

That the Clebsch-Gordan coefficients may be taken as real is the so-called "Condon-Shortley convention." It follows from the fact that the maximum values of j and m are $j_{\max} = j_1 + j_2 = m_{\max}$ so that (leaving out the fixed j_1 and j_2 labels of the two particles so as not to be too confusing)

$$\begin{aligned}|j_{\max} m_{\max}\rangle &= \sum_{m_1+m_2=m_{\max}} |m_1 m_2\rangle\langle m_1 m_2|j_{\max} m_{\max}\rangle \\ &= |j_1 j_2\rangle\langle j_1 j_2|j_{\max} m_{\max}\rangle.\end{aligned}$$

But then

$$\langle j_{\max} m_{\max}|j_{\max} m_{\max}\rangle = \langle j_{\max} m_{\max}|j_1 j_2\rangle\langle j_1 j_2|j_1 j_2\rangle\langle j_1 j_2|j_{\max} m_{\max}\rangle$$

or (since $\langle jm|jm\rangle = \langle m_1 m_2|m_1 m_2\rangle = 1$)

$$1 = |\langle j_{\max} m_{\max}|j_1 j_2\rangle|^2 .$$

Therefore we take as our (Condon-Shortley) convention $\langle j_{\max} m_{\max}|j_1 j_2\rangle := +1$, and this forces the rest of the Clebsch-Gordan coefficients to be real also since they are constructed by repeated application of the operator $J_- = J_{1-} + J_{2-}$ to the $|jm\rangle$ and $|m_1 m_2\rangle$ states starting at the top.

Let us see what we can say about the rotation operator that acts on composite angular momentum states. The corresponding rotation operator is

$$e^{-i\boldsymbol{\theta}\cdot\mathbf{J}} = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_1} \otimes e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_2} = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_1} e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_2}$$

where the last equality is just a commonly used shorthand notation. Writing the rotation operator as the direct product $U(R) = U_1(R) \otimes U_2(R)$, its action is defined by

$$\begin{aligned}[U_1(R) \otimes U_2(R)]|j_1 j_2 m_1 m_2\rangle &= [U_1(R) \otimes U_2(R)](|j_1 m_1\rangle \otimes |j_2 m_2\rangle) \\ &= U_1(R)|j_1 m_1\rangle \otimes U_2(R)|j_2 m_2\rangle.\end{aligned}$$

Applying equation (100) we can now write

$$\begin{aligned}[U_1(R) \otimes U_2(R)]|j_1 j_2 m_1 m_2\rangle &= U_1(R)|j_1 m_1\rangle \otimes U_2(R)|j_2 m_2\rangle \\ &= \sum_{m_1' m_2'} |j_1 m_1'\rangle \otimes |j_2 m_2'\rangle \mathscr{D}^{(j_1)}_{m_1' m_1}(\boldsymbol{\theta}) \mathscr{D}^{(j_2)}_{m_2' m_2}(\boldsymbol{\theta}) \\ &= \sum_{m_1' m_2'} |j_1 j_2 m_1' m_2'\rangle \mathscr{D}^{(j_1)}_{m_1' m_1}(\boldsymbol{\theta}) \mathscr{D}^{(j_2)}_{m_2' m_2}(\boldsymbol{\theta}) \qquad (108)\end{aligned}$$

which implies

$$\langle j_1 j_2 m_1' m_2'|U_1(R) \otimes U_2(R)|j_1 j_2 m_1 m_2\rangle = \mathscr{D}^{(j_1)}_{m_1' m_1}(\boldsymbol{\theta}) \mathscr{D}^{(j_2)}_{m_2' m_2}(\boldsymbol{\theta}). \qquad (109)$$

We regard the left side of this equation as the $m_1'm_2', m_1m_2$ matrix element of a $[(2j_1+1)(2j_2+1) \times (2j_1+1)(2j_2+1)]$-dimensional matrix called the **direct** (or **Kronecker**) **product** of $\mathscr{D}^{(j_1)}(\boldsymbol{\theta})$ and $\mathscr{D}^{(j_2)}(\boldsymbol{\theta})$ and written $\mathscr{D}^{(j_1)}(\boldsymbol{\theta}) \otimes \mathscr{D}^{(j_2)}(\boldsymbol{\theta})$. In other words

$$[\mathscr{D}^{(j_1)}(\boldsymbol{\theta}) \otimes \mathscr{D}^{(j_2)}(\boldsymbol{\theta})]_{m_1'm_2', m_1m_2} = \mathscr{D}^{(j_1)}_{m_1'm_1}(\boldsymbol{\theta})\mathscr{D}^{(j_2)}_{m_2'm_2}(\boldsymbol{\theta}).$$

The double subscripts make the direct product of two matrices look confusing, but the result really isn't if we just use the definitions carefully. So to understand this, we need to take another small linear algebra detour.

If we have a vector space V with basis $\{e_1, \dots, e_n\}$ and another space V' with basis $\{e_1', \dots, e_m'\}$, then we have the space $V \otimes V'$ with the basis $\{e_i \otimes e_j'\}$ which we take to have the ordering

$$\{e_1 \otimes e_1', e_1 \otimes e_2', \dots, e_1 \otimes e_m', e_2 \otimes e_1', \dots, e_2 \otimes e_m', \dots, e_n \otimes e_1', \dots, e_n \otimes e_m'\}.$$

Given the operators $A \in L(V)$ and $B \in L(V')$, the matrix representations of A and B are defined by $Ae_i = \sum_k e_k a_{ki}$ and $Be_j' = \sum_l e_l' b_{lj}$, and thus the matrix representation of $A \otimes B$ is also defined in the same way by

$$(A \otimes B)(e_i \otimes e_j') = Ae_i \otimes Be_j' = \sum_{kl}(e_k \otimes e_l')a_{ki}b_{lj} = \sum_{kl}(e_k \otimes e_l')(A \otimes B)_{kl,ij}.$$

But Ae_i is just the ith column of the matrix (a_{ij}), and now we have each row and column labeled by a double subscript so that, e.g., the $(1,1)$th column of the matrix representation of $A \otimes B$ is given relative to the above ordered basis by

$$\begin{aligned}
&(A \otimes B)(e_1 \otimes e_1') \\
&\qquad = \sum_{kl}(e_k \otimes e_l')a_{k1}b_{l1} \\
&\qquad = \{a_{11}b_{11}, a_{11}b_{21}, \dots, a_{11}b_{m1}, a_{21}b_{11}, \dots, a_{21}b_{m1}, \dots, a_{n1}b_{11}, \dots, a_{n1}b_{m1}\}
\end{aligned}$$

Since this is really a column vector, we see by careful inspection that it is just a_{11} times the first column of B followed by a_{21} times the first column of B, and so on down to a_{n1} times the first column of B. And in general, the matrix of $A \otimes B$ is given in block matrix form by

$$\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{bmatrix}.$$

There is another basic result from linear algebra that we should review before treating $U(R)$. Suppose we have an operator $T \in L(V)$, and let W be an invariant subspace of V. By this we mean that $T(W) \subset W$. Let $\{e_1, \dots, e_m\}$

be a basis for W, and extend this to a basis $\{e_1, \dots, e_m, e_{m+1}, \dots, e_n\}$ for V. By definition of invariant subspace, for any $1 \le i \le m$ the effect of T on e_i is given by

$$Te_i = \sum_{j=1}^{m} e_j a_{ji} \qquad 1 \le i \le m$$

for appropriate scalars a_{ji}. Compare this to

$$Te_i = \sum_{j=1}^{m} e_j b_{ji} + \sum_{j=m+1}^{n} e_j c_{ji} \qquad m+1 \le i \le n$$

for some scalars b_{ji} and c_{ji} where we have assumed that the subspace spanned by $\{e_{m+1}, \dots, e_n\}$ is not an invariant subspace itself. Since the ith column of $[T] = (a_{ij})$ is given by Te_i, a moments thought should convince you that the matrix representation for T will be of the block matrix form

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

If it turns out that $\{e_{m+1}, \dots, e_n\}$ also spans an invariant subspace W', then for $m+1 \le i \le n$ we will have $Te_i = \sum_{j=m+1}^{n} e_j c_{ji}$ and the representation of T will look like

$$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}.$$

In this case we have $V = W \oplus W'$ (the direct sum) and we write $T = T_W \oplus T_{W'}$ where T_W is the restriction of T to W and $T_{W'}$ is the restriction of T to W'. (As a reminder, we say that V is the **direct sum** of subspaces $W_1, \dots, W_r$ if the union of bases for the W_i is a basis for V, and in this case we write $V = W_1 \oplus \cdots \oplus W_r$. If each of these W_i is also an invariant subspace with respect to some operator T, then we write $T = T_{W_1} \oplus \cdots \oplus T_{W_r}$ and its matrix representation with respect to this basis will be block diagonal.)

Now let's go back to the rotation operator $U(R) = U_1(R) \otimes U_2(R)$. When we say that a representation is reducible, we mean there exists a basis for the state space relative to which the matrix representation of *all* relevant operators is block diagonal. We now want to show that the set of direct product matrices is reducible, and thus in the $|j_1 j_2 j m\rangle$ basis the matrix representation of the direct product takes the form

$$\begin{bmatrix} \square & & & \\ & \square & & \\ & & \ddots & \\ & & & \square \end{bmatrix}$$

where each block represents the $(2j+1)\times(2j+1)$ matrix $\mathscr{D}^{(j)}(\boldsymbol{\theta})$ for each value of $j = j_1 + j_2, j_1 + j_2 - 1, \ldots, |j_1 - j_2|$. Symbolically we write this as

$$\mathscr{D}^{(j_1)}(\boldsymbol{\theta}) \otimes \mathscr{D}^{(j_2)}(\boldsymbol{\theta}) = \mathscr{D}^{(j_1+j_2)}(\boldsymbol{\theta}) \oplus \mathscr{D}^{(j_1+j_2-1)}(\boldsymbol{\theta}) \oplus \cdots \oplus \mathscr{D}^{|j_1-j_2|}(\boldsymbol{\theta}) \tag{110}$$

The idea here is just a somewhat complicated version of the usual result from linear algebra dealing with the matrix representation of a linear operator under a change of basis. So let's first review that topic.

Given an operator T, its representation $A = (a_{ij}) := [T]_e$ on a basis $\{e_i\}$ is defined by $Te_i = \sum_j e_j a_{ji}$. If we want to change to a new basis $\{e'_i\}$, then we first define the transition matrix P by $e'_i = \sum_j e_j p_{ji}$. Then on the one hand we have the representation $A' = (a'_{ij}) := [T]_{e'}$ given by

$$Te'_i = \sum_j e'_j a'_{ji} = \sum_j \sum_k e_k p_{kj} a'_{ji}$$

while on the other hand we have

$$Te'_i = T\Big(\sum_j e_j p_{ji}\Big) = \sum_j (Te_j) p_{ji} = \sum_j \sum_k e_k a_{kj} p_{ji}.$$

Equating these last two results and using the fact that the $\{e_k\}$ are a basis (and hence linearly independent) we have

$$\sum_j p_{kj} a'_{ji} = \sum_j a_{kj} p_{ji}$$

which in matrix notation is just $PA' = AP$. The transition matrix P must be nonsingular because either basis can be written in terms of the other, so we are left with the fundamental result

$$A' = P^{-1} AP.$$

Furthermore, if both bases are orthonormal, then P will be unitary so that $P^{-1} = P^\dagger = P^{*T}$ and we have $A' = P^\dagger AP$. In terms of components this is

$$a'_{ij} = \sum_{kl} p^*_{ki}\, a_{kl}\, p_{lj}. \tag{111}$$

Now let's go back and prove equation (110). We first insert complete sets of the states $|jm\rangle$ into the left side of equation (109) to obtain (again leaving out the j_1 and j_2 for neatness)

$$\begin{aligned}
\langle m'_1 m'_2|U(R)|m_1 m_2\rangle &= \sum_{\substack{j'm'\\ jm}} \langle m'_1 m'_2|j'm'\rangle\langle j'm'|U(R)|jm\rangle\langle jm|m_1 m_2\rangle \\
&= \sum_{\substack{j'm'\\ jm}} \langle m'_1 m'_2|j'm'\rangle\langle m_1 m_2|jm\rangle \mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta})\delta_{j'j} \\
&= \sum_{j'm'm} \langle m'_1 m'_2|j'm'\rangle\langle m_1 m_2|j'm\rangle \mathscr{D}^{(j')}_{m'm}(\boldsymbol{\theta}) \qquad (112)
\end{aligned}$$

where the second line follows from equations (98), (101) and the fact that the Clebsch-Gordan coefficients are real. From equation (109) we then have (after relabeling the dummy index $j' \to j$)

$$\mathscr{D}^{(j_1)}_{m'_1 m_1}(\boldsymbol{\theta})\mathscr{D}^{(j_2)}_{m'_2 m_2}(\boldsymbol{\theta}) = \sum_{jm'm} \langle m'_1 m'_2 | jm'\rangle\langle m_1 m_2 | jm\rangle \mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta}) \tag{113}$$

where the sum is over $|j_1 - j_2| \le j \le j_1 + j_2$. Equation (113) is sometimes called the **Clebsch-Gordan series**. Also note that there is no real sum over m' and m in this equation because we must have $m' = m'_1 + m'_2$ and $m = m_1 + m_2$.

Analogous to the a_{kl} in equation (111), we write equation (112) as

$$\begin{aligned}\langle m'_1 m'_2 | U_1(R) \otimes U_2(R) | m_1 m_2\rangle &:= R_{m'_1 m'_2, m_1 m_2} \\ &= \sum_{j'm'm} \langle m'_1 m'_2 | j'm'\rangle\langle m_1 m_2 | j'm\rangle \mathscr{D}^{(j')}_{m'm}(\boldsymbol{\theta}).\end{aligned}$$

From equation (106) we have the (real) transition matrix elements which we write as

$$p_{m_1 m_2, jm} = \langle m_1 m_2 | jm\rangle.$$

Defining the analogue of a'_{ij} in equation (111) by

$$R'_{j'm', jm} := \langle j'm' | U(R) | jm\rangle$$

we then have

$$\begin{aligned}\langle j'm' | U(R) | jm\rangle &= \sum_{\substack{\bar m_1 \bar m_2 \\ \bar m'_1 \bar m'_2}} \langle \bar m_1 \bar m_2 | j'm'\rangle R_{\bar m_1 \bar m_2, \bar m'_1 \bar m'_2} \langle \bar m'_1 \bar m'_2 | jm\rangle \\ &= \sum_{\substack{\bar m_1 \bar m_2 \\ \bar m'_1 \bar m'_2}} \sum_{j'' m'' \bar m''} \langle \bar m_1 \bar m_2 | j'm'\rangle\langle \bar m_1 \bar m_2 | j''m''\rangle\langle \bar m'_1 \bar m'_2 | j'' \bar m''\rangle \\ &\qquad\qquad \times \mathscr{D}^{(j'')}_{m'' \bar m''}(\boldsymbol{\theta})\langle \bar m'_1 \bar m'_2 | jm\rangle.\end{aligned}$$

But

$$\sum_{\bar m_1 \bar m_2} \langle j'm' | \bar m_1 \bar m_2\rangle\langle \bar m_1 \bar m_2 | j''m''\rangle = \langle j'm' | j''m''\rangle = \delta_{j'j''}\delta_{m'm''}$$

and

$$\sum_{\bar m'_1 \bar m'_2} \langle j'' \bar m'' | \bar m'_1 \bar m'_2\rangle\langle \bar m'_1 \bar m'_2 | jm\rangle = \langle j'' \bar m'' | jm\rangle = \delta_{j''j}\delta_{\bar m'' m}$$

so we are left with

$$\langle j'm' | U(R) | jm\rangle = \sum_{j'' m'' \bar m''} \delta_{j'j''}\delta_{m'm''}\delta_{j''j}\delta_{\bar m'' m}\mathscr{D}^{(j'')}_{m'' \bar m''}(\boldsymbol{\theta}) = \mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta})\delta_{j'j} \tag{114}$$

If on the two particle angular momentum space we take our *ordered* basis $\{|jm\rangle\}$ to be

$$\{|j_1+j_2, j_1+j_2\rangle, |j_1+j_2, j_1+j_2-1\rangle, \dots, |j_1+j_2, -(j_1+j_2)\rangle,$$
$$|j_1+j_2-1, j_1+j_2-1\rangle, \dots, |j_1+j_2-1, -(j_1+j_2-1)\rangle,$$
$$\dots, |\,|j_1-j_2|\,, |j_1-j_2|\rangle, \dots, |\,|j_1-j_2|\,, -|j_1-j_2|\rangle\}$$

then equation (114) is just equation (110) in terms of components.

Of course, had we just started with $U(R) = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}}$ and used equations (99) and (100) we would have again had

$$\langle j'm'|U(R)|jm\rangle = \sum_{m''}\langle j'm'|jm''\rangle \mathscr{D}^{(j)}_{m''m}(\boldsymbol{\theta}) = \mathscr{D}^{(j)}_{m'm}(\boldsymbol{\theta})\delta_{j'j}.$$

But this doesn't show that in fact the matrix (112) can be put into the block diagonal form (110). In other words, the operator $e^{-i\boldsymbol{\theta}\cdot\mathbf{J}}$ acts on the states $|jm\rangle$ in a manifestly block diagonal manner (due to equation (100)), whereas the operator $e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_1} \otimes e^{-i\boldsymbol{\theta}\cdot\mathbf{J}_2}$ acts on the states $|m_1 m_2\rangle$ in the complicated manner of equation (108), and we had to show that this could in fact be put into block diagonal form by the appropriate change of basis.

10 Lorentz Invariance and Spin

Before turning to another derivation of the Dirac equation, we need to take a more careful look at the **inhomogeneous Lorentz group**, also called the **Poincaré group**. These are transformations of the form

$$\overline{x}^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu \tag{115}$$

where we are using the metric

$$g_{\mu\nu} = g^{\mu\nu} = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix}$$

and Λ satisfies

$$\Lambda^T g \Lambda = g \qquad \text{or} \qquad g_{\mu\nu}\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta = g_{\alpha\beta}. \tag{116}$$

(This is a consequence of requiring that $\overline{x}^\mu \overline{x}_\mu = x^\nu x_\nu$ where $\overline{x}^\mu = \Lambda^\mu{}_\nu x^\nu$.) We will refer to the transformation (115) as simply a "Lorentz transformation," and sometimes abbreviate it by simply LT. If we compose two successive transformations (115) we obtain

$$\begin{aligned}\overline{\overline{x}}^\mu &= \overline{\Lambda}^\mu{}_\nu \overline{x}^\nu + \overline{a}^\mu \\ &= \overline{\Lambda}^\mu{}_\nu(\Lambda^\nu{}_\alpha x^\alpha + a^\nu) + \overline{a}^\mu \\ &= (\overline{\Lambda}^\mu{}_\nu \Lambda^\nu{}_\alpha)x^\alpha + (\overline{\Lambda}^\mu{}_\nu a^\nu + \overline{a}^\mu)\end{aligned} \tag{117}$$

which is just another LT, and hence the set of transformations (115) forms a group called the "Poincaré group" as claimed.

From $\Lambda^T g \Lambda = g$ we see that $(\det \Lambda)^2 = 1$ so that $\det \Lambda = \pm 1$. Since the identity transformation has determinant equal to $+1$, we restrict ourselves to the case $\det \Lambda = +1$ because we will want to look at transformations that differ infinitesimally from the identity transformation. It also follows from equation (116) that $g_{\mu\nu}\Lambda^\mu{}_0\Lambda^\nu{}_0 = 1$ or $(\Lambda^0{}_0)^2 = 1 + \sum_i (\Lambda^i{}_0)^2 \geq 1$ so that

$$\Lambda^0{}_0 \geq 1 \qquad \text{or} \qquad \Lambda^0{}_0 \leq -1.$$

Since only the case $\Lambda^0{}_0 \geq 1$ can be used for an infinitesimal transformation, we will restrict ourselves from now on to the set of **proper, orthochronous Lorentz transformations**. However, note that the set of all Lorentz transformations can be divided into four classes given by $\det \Lambda = \pm 1$ and either $\Lambda^0{}_0 \geq 1$ or $\Lambda^0{}_0 \leq -1$.

Our goal is to relate this to quantum mechanics and the description of quantum mechanical states, so to each transformation (115) we shall associate a unitary operator on the (infinite-dimensional) Hilbert space of all physical state vectors, denoted by $U(a, \Lambda)$. The effect of this operator is defined by

$$U(a, \Lambda)\psi(x) := \psi(\Lambda x + a).$$

From equation (115) we have

$$U(a, \Lambda) = U(a, 1)U(0, \Lambda) := U(a)U(\Lambda). \tag{118}$$

Then using equation (117) we have the multiplication law

$$U(a', \Lambda')U(a, \Lambda) = U(\Lambda' a + a', \Lambda'\Lambda) \tag{119}$$

and therefore, in particular

$$U(a')U(a) = U(a' + a) \tag{120a}$$

$$U(\Lambda')U(\Lambda) = U(\Lambda'\Lambda). \tag{120b}$$

For an infinitesimal transformation

$$\Lambda^\mu{}_\nu = g^\mu{}_\nu + \omega^\mu{}_\nu \tag{121}$$

we have from equation (116)

$$\begin{aligned} g_{\alpha\beta} &= (g^\mu{}_\alpha + \omega^\mu{}_\alpha)(g^\nu{}_\beta + \omega^\nu{}_\beta)g_{\mu\nu} = (g^\mu{}_\alpha + \omega^\mu{}_\alpha)(g_{\mu\beta} + \omega_{\mu\beta}) \\ &= g_{\alpha\beta} + \omega_{\beta\alpha} + \omega_{\alpha\beta} \end{aligned}$$

and therefore

$$\omega_{\alpha\beta} = -\omega_{\beta\alpha}.$$

Then we write our operators in the form

$$U(a) = e^{ia_\mu P^\mu} \tag{122a}$$

$$U(\Lambda) = e^{\frac{i}{2}\omega_{\mu\nu} M^{\mu\nu}} \tag{122b}$$

where the generators P^μ and $M^{\mu\nu}$ remain to be specified. (This is just exponentiating the general form of an infinitesimal transformation. Note also that we have included a factor of i in the definition of $U(\Lambda)$ (as compared to equation (92)) because we are now talking about quantum mechanical operators.)

Before proceeding, let us define just what is meant when we say that an operator transforms as a vector. Recall from linear algebra (yet again) that under a rotation $R : \mathbf{v} \to \mathbf{v}'$ we have $R\mathbf{v} = R(v^i e_i) = v^i Re_i = v^i e'_j R^j{}_i$. But $R\mathbf{v} = \mathbf{v}' = v'^j e'_j$ so that $v'^j = R^j{}_i v^i$ defines how the components of a vector transform. In quantum mechanics we require that the expectation values of an operator transform in the same way. Thus, if $|\psi\rangle \to |\psi'\rangle = U(R)|\psi\rangle$ where $U(R)$ is the unitary quantum mechanical operator corresponding to the rotation R, then an operator $\mathbf{A}$ is defined to be a **vector operator** if

$$\langle\psi'|A^i|\psi'\rangle = R^i{}_j \langle\psi|A^j|\psi\rangle.$$

Using $|\psi'\rangle = U(R)|\psi\rangle$ and the fact that $|\psi\rangle$ is arbitary, this yields

$$U^\dagger(R) A^i U(R) = R^i{}_j A^j \tag{123}$$

where $(R^i{}_j)$ is a real, orthogonal matrix that represents rotations in $\mathbb{R}^3$.

If we require that the expectation value of a rotated operator in the original state be the same as the expectation value of the original operator in the rotated state, then $\langle\psi|\mathbf{A}'|\psi\rangle = \langle\psi'|\mathbf{A}|\psi'\rangle = \langle\psi|U^\dagger(R)\mathbf{A}U(R)|\psi\rangle$ so that

$$A'^i = U^\dagger(R) A^i U(R) = R^i{}_j A^j$$

which shows the meaning of equation (123). (This is just the difference between active and passive transformations.) In addition, we can multiply both sides of equation (123) by $(R^{-1})^k{}_i$ and sum to write $U^\dagger(R)(R^{-1})^k{}_i A^i U(R) = A^k$ so that (since $U^\dagger = U^{-1}$)

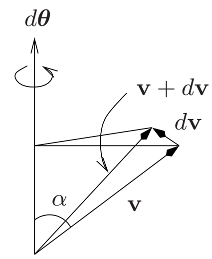
$$U(R) A^k U^{-1}(R) = (R^{-1})^k{}_i A^i \tag{124}$$

which we take as an equivalent definition of a vector operator.

The only difference between this rotational case and our present purposes is that now we are using a Lorentz transformation $\Lambda^\mu{}_\nu$ instead of the rotation matrix $R^i{}_j$, but otherwise everything is exactly the same since Λ is also an orthogonal transformation. (In fact, this is the same as we saw earlier with equation (31) describing why the Dirac gamma matrices are said to transform as vectors.)

(It should be pointed out that what we have done is *not* the same as changing between "pictures" such as the Schrödinger and Heisenberg representations. In that case, we require that the expectation value of an operator in the Schrödinger picture in a Schrödinger state be the same as the expectation value of a Heisenberg operator in a Heisenberg state. In other words, $\langle\psi_S|A_S|\psi_S\rangle := \langle\psi_H|A_H|\psi_H\rangle$, and if $|\psi_H\rangle = U(t)|\psi_S\rangle (= e^{+iEt}|\psi_S\rangle)$, then $\langle\psi_H|A_H|\psi_H\rangle = \langle\psi_S|U^\dagger(t)A_H U(t)|\psi_S\rangle$ so that $A_H = U(t)A_S U^\dagger(t)$. Note that the $U(t)$ and $U^\dagger(t)$ are reversed from what they are in equation (123).)

Another point that we should discuss is the commutation relation between vector operators and the generator of rotations, i.e., the angular momentum operators. These will serve to help classify certain operators. So, recall from classical mechanics that a vector $\mathbf{v}$ undergoing a rotation $d\boldsymbol{\theta}$ changes by an amount $d\mathbf{v}$ as shown in the figure below.



We have $\|d\mathbf{v}\| = \|\mathbf{v}\| \sin\alpha \, \|d\boldsymbol{\theta}\|$ in the direction shown (i.e., perpendicular to the plane defined by $\mathbf{v}$ and $d\boldsymbol{\theta}$) so that $d\mathbf{v} = d\boldsymbol{\theta} \times \mathbf{v}$ and $\mathbf{v} \to \mathbf{v} + d\mathbf{v}$.

But we just saw above that under a rotation R we have $v^i \to R^i{}_j v^j$ so that (using $(d\boldsymbol{\theta} \times \mathbf{v})^i = \varepsilon^{ijk} d\theta_j v_k$ and then relabeling)

$$v^i \to v'^i = R^i{}_j v^j = v^i + dv^i = v^i + \varepsilon^{ijk} d\theta_j v_k = (\delta^{ij} + \varepsilon^{ikj} d\theta_k) v_j$$

which shows that to first order we have (where in this case there is no difference between upper and lower indices since in $\mathbb{R}^3$ with cartesian coordinates we have $g_{ij} = \delta_{ij}$ anyway)

$$R^i{}_j = \delta^i{}_j - \varepsilon^i{}_{jk} d\theta^k. \tag{125}$$

Next, expand the left side of (123) to first order using $U(R) = e^{-id\boldsymbol{\theta}\cdot\mathbf{J}}$ to yield

$$(1 + id\boldsymbol{\theta}\cdot\mathbf{J})A^i(1 - id\boldsymbol{\theta}\cdot\mathbf{J}) = A^i - id\theta_j[A^i, J^j].$$

Finally, using equation (125) in (123) we obtain $[A^i, J^j] = i\varepsilon^{ijk} A_k$ which we can write as

$$[J^i, A^j] = i\varepsilon^{ijk} A_k. \tag{126}$$

This then is the commutation relation of a vector operator with the generator of angular momentum.

Now back to where we were. To define the Poincaré algebra, we need to obtain the commutation relations for the generators. Since $[A, B] = 0$ implies $e^A e^B = e^{A+B}$, we see from equations (120a) and (122a) that

$$[P^\mu, P^\nu] = 0. \tag{127}$$

Now put equation (118) into (119) and use (120b):

$$U(a')U(\Lambda')U(a)U(\Lambda) = U(a')U(\Lambda' a)U(\Lambda')U(\Lambda)$$

so that

$$U(\Lambda')U(a)U^{-1}(\Lambda') = U(\Lambda' a)$$

or (dropping the prime on Λ)

$$U(\Lambda)e^{ia_\mu P^\mu}U^{-1}(\Lambda) = e^{i\Lambda_\mu{}^\nu a_\nu P^\mu}.$$

To first order this is

$$U(\Lambda)(1 + ia_\mu P^\mu)U^{-1}(\Lambda) = 1 + iP^\nu \Lambda_\nu{}^\mu a_\mu$$

and hence

$$U(\Lambda)P^\mu U^{-1}(\Lambda) = P^\nu \Lambda_\nu{}^\mu = (\Lambda^{-1})^\mu{}_\nu P^\nu \tag{128}$$

so that comparison with equation (124) shows that the operator P^μ transforms as a 4-vector.

An equivalent way of writing equation (128) is

$$U^{-1}(\Lambda)P^\mu U(\Lambda) = \Lambda^\mu{}_\nu P^\nu. \tag{129}$$

It is worth pointing out the physical meaning of this equation. Suppose we have an eigenstate of P^μ:

$$P^\mu|\Psi_p\rangle = p^\mu|\Psi_p\rangle.$$

Then acting on $|\Psi_p\rangle$ with $U(\Lambda)$ results in a state with eigenvalue $p'^\mu = \Lambda^\mu{}_\nu p^\nu$ because

$$\begin{aligned} P^\mu[U(\Lambda)|\Psi_p\rangle] &= U(\Lambda)[U^{-1}(\Lambda)P^\mu U(\Lambda)]|\Psi_p\rangle = U(\Lambda)[\Lambda^\mu{}_\nu p^\nu]|\Psi_p\rangle \\ &= \Lambda^\mu{}_\nu p^\nu[U(\Lambda)|\Psi_p\rangle]. \end{aligned} \tag{130}$$

In other words, $U(\Lambda)$ boosts a state with eigenvalue p to a state with eigenvalue Λp. Alternatively, we can say that the expectation value of the operator P^μ in the boosted state is the same as the expectation value of the boosted operator ΛP in the original state.

Next we want to find the commutation relations for the generators $M^{\mu\nu}$. Using equation (122b) for infinitesimal $\omega_{\mu\nu}$ together with equation (121) in (128) we have

$$(1 + \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu})P^\alpha(1 - \frac{i}{2}\omega_{\rho\sigma}M^{\rho\sigma}) = (g^\alpha{}_\beta - \omega^\alpha{}_\beta)P^\beta$$

or

$$(1 + \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu})(P^\alpha - \frac{i}{2}\omega_{\rho\sigma}P^\alpha M^{\rho\sigma}) = P^\alpha - \omega^\alpha{}_\beta P^\beta$$

so that (to first order as usual)

$$\begin{aligned} P^\alpha - \frac{i}{2}\omega_{\mu\nu}P^\alpha M^{\mu\nu} + \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu}P^\alpha &= P^\alpha - \omega^\alpha{}_\beta P^\beta = P^\alpha - g^{\alpha\mu}\omega_{\mu\nu}P^\nu \\ &= P^\alpha - \frac{1}{2}\omega_{\mu\nu}(g^{\alpha\mu}P^\nu - g^{\alpha\nu}P^\mu) \end{aligned}$$

where in the last line we antisymmetrized over the indices μ and ν. Canceling P^α and equating the coefficients of $\omega_{\mu\nu}$ (which is just some arbitrary number) we are left with

$$[M^{\mu\nu}, P^\alpha] = -i(P^\mu g^{\nu\alpha} - P^\nu g^{\mu\alpha}). \tag{131}$$

It will be important to us later to realize that this result applies to any 4-vector, and not just P^μ. This is because equation (128) applies to any 4-vector by definition.

Since $\omega_{\mu\nu}$ is antisymmetric, we may just as well assume that $M^{\mu\nu}$ is too, so let us define the quantities (essentially the same as we did in Section 8)

$$K_i = M_{i0} \qquad \text{and} \qquad J_i = \frac{1}{2}\varepsilon_{ijk}M^{jk}. \tag{132}$$

In order to find the commutation relations for the J_i and K_i, we need those for the $M^{\mu\nu}$. To obtain these, we perform the following manipulations.

First note that

$$U(\Lambda)U^{-1}(\Lambda) = 1 = U(\Lambda\Lambda^{-1}) = U(\Lambda)U(\Lambda^{-1})$$

and therefore

$$U^{-1}(\Lambda) = U(\Lambda^{-1}).$$

(This is really just a special case of the general result from group theory that if $\varphi : G \to G'$ is a homomorphism of groups, then for any $g \in G$ we have $e' = \varphi(e) = \varphi(gg^{-1}) = \varphi(g)\varphi(g^{-1})$ so that $\varphi(g)^{-1} = \varphi(g^{-1})$. In this case, the operator $U(\Lambda)$ is the image of the group element Λ under the homomorphism $U : \Lambda \to U(\Lambda)$. We are also writing $U^{-1}(\Lambda)$ rather than the more correct $U(\Lambda)^{-1}$, but this is common practice.) Next we have

$$U(\Lambda)U(\Lambda')U^{-1}(\Lambda) = U(\Lambda\Lambda'\Lambda^{-1})$$

which we want to evaluate for infinitesimal transformations. In a symbolic notation, if we have $\Lambda' = 1 + \omega'$ so that $\Lambda\Lambda'\Lambda^{-1} = 1 + \Lambda\omega'\Lambda^{-1} := 1 + \widetilde{\omega}$, then

$$\widetilde{\omega}_{\mu\nu} = \Lambda_\mu{}^\alpha\omega'_{\alpha\beta}(\Lambda^{-1})^\beta{}_\nu = (\Lambda^{-1})^\alpha{}_\mu(\Lambda^{-1})^\beta{}_\nu\omega'_{\alpha\beta}.$$

Therefore

$$U(\Lambda\Lambda'\Lambda^{-1}) = 1 + \frac{i}{2}\widetilde{\omega}_{\mu\nu}M^{\mu\nu} = 1 + \frac{i}{2}(\Lambda^{-1})^\alpha{}_\mu(\Lambda^{-1})^\beta{}_\nu\omega'_{\alpha\beta}M^{\mu\nu}. \tag{133}$$

But

$$\begin{aligned} U(\Lambda)U(\Lambda')U^{-1}(\Lambda) &= U(\Lambda)(1 + \frac{i}{2}\omega'_{\alpha\beta}M^{\alpha\beta})U^{-1}(\Lambda) \\ &= 1 + \frac{i}{2}\omega'_{\alpha\beta}U(\Lambda)M^{\alpha\beta}U^{-1}(\Lambda) \end{aligned} \tag{134}$$

so that equating equations (133) and (134) yields

$$U(\Lambda)M^{\alpha\beta}U^{-1}(\Lambda) = (\Lambda^{-1})^\alpha{}_\mu(\Lambda^{-1})^\beta{}_\nu M^{\mu\nu}. \tag{135}$$

This equation shows (by definition) that $M^{\mu\nu}$ transforms as a second rank tensor operator.

Using $(\Lambda^{-1})^{\alpha}{}_{\mu} = g^{\alpha}{}_{\mu} - \omega^{\alpha}{}_{\mu}$ in equation (135) we have

$$(1 + \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu})M^{\alpha\beta}(1 - \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu}) = (g^{\alpha}{}_{\mu} - \omega^{\alpha}{}_{\mu})(g^{\beta}{}_{\nu} - \omega^{\beta}{}_{\nu})M^{\mu\nu}$$

which to first order expands to

$$M^{\alpha\beta} - \frac{i}{2}\omega_{\mu\nu}M^{\alpha\beta}M^{\mu\nu} + \frac{i}{2}\omega_{\mu\nu}M^{\mu\nu}M^{\alpha\beta} = M^{\alpha\beta} - \omega^{\beta}{}_{\nu}M^{\alpha\nu} - \omega^{\alpha}{}_{\mu}M^{\mu\beta}$$

or

$$\frac{i}{2}\omega_{\mu\nu}[M^{\mu\nu}, M^{\alpha\beta}] = -M^{\alpha\nu}\omega^{\beta}{}_{\nu} - M^{\nu\beta}\omega^{\alpha}{}_{\nu}.$$

Antisymmetrizing over μ and ν we can write the two terms on the right side of this equation as

$$M^{\alpha\nu}\omega^{\beta}{}_{\nu} = M^{\alpha\nu}\omega_{\mu\nu}g^{\mu\beta} = \frac{1}{2}\omega_{\mu\nu}(M^{\alpha\nu}g^{\mu\beta} - M^{\alpha\mu}g^{\nu\beta})$$

and

$$M^{\nu\beta}\omega^{\alpha}{}_{\nu} = M^{\nu\beta}\omega_{\mu\nu}g^{\mu\alpha} = \frac{1}{2}\omega_{\mu\nu}(M^{\nu\beta}g^{\mu\alpha} - M^{\mu\beta}g^{\nu\alpha}).$$

Using these, we finally obtain the commutation relation for the generators $M^{\mu\nu}$

$$[M^{\mu\nu}, M^{\alpha\beta}] = i(M^{\alpha\nu}g^{\mu\beta} + M^{\nu\beta}g^{\mu\alpha} - M^{\alpha\mu}g^{\nu\beta} - M^{\mu\beta}g^{\nu\alpha}). \tag{136}$$

Equations (127), (131) and (136) define the **Poincaré algebra**. An explicit example of the generators is provided by the matrix $\Sigma^{\mu\nu}$ defined in equation (46). You should be able to see the reason for this.

(Technically, one says that these generators define the Lie algebra of the Poincaré group, so let me briefly explain what is going on. The Poincaré (or inhomogeneous Lorentz) group consists of translations (in four directions), rotations (with three degrees of freedom) and boosts (in three directions), each of which is specified by some set of parameters. In other words, we can think of each group element as a point in a (in this case) 10-dimensional space. To each point in this space we associate an operator $U(a, \Lambda) = U(a)U(\Lambda)$ and hence a matrix. We expand these operators in a neighborhood of the identity, and as each group parameter is varied, a curve is traced out in the differentiable manifold that defines the group space. The derivative of this coordinate curve with respect to its parameter is the tangent vector to the curve at that point. The generators are then the tangent vectors to each coordinate curve at the identity, and the vector space spanned by the generators is the **Lie algebra** (which is then the tangent space at the identity). (In general, an **algebra** is a vector space on which we have also defined the vector product of two vectors. In this case, the product is the commutator (or **Lie bracket**).) In other words, the generators are a basis for the tangent space at the identity (the Lie algebra). It can be shown that by exponentiating the Lie algebra, one can construct every group element (i.e., the operators $U(a,\Lambda)$) at least in a neighborhood of the identity. The question of whether or not we can get to *every* group element by exponentiating the Lie algebra is a nontrivial one, and the answer depends on the topology of the group in question.)

At last we are ready to find the commutation relations for the generators $\mathbf{J}$ and $\mathbf{K}$. Remember that now we are doing this calculation in $\mathbb{R}^3$ with the metric $g_{ij} = \delta_{ij}$ as discussed following equations (93). Furthermore, ε_{ijk} is just the permutation *symbol* so it doesn't carry spacetime (tensor) indices. Then using

$$\varepsilon_{ijk}\varepsilon^{klm} = \delta_i^l \delta_j^m - \delta_i^m \delta_j^l$$

or the equivalent form

$$\varepsilon_{ijk}\varepsilon_{klm} = \delta_{li}\delta_{mj} - \delta_{mi}\delta_{lj}$$

we have

$$\begin{aligned}
[J_i, J_j] &= \frac{1}{4}\varepsilon_{ilm}\varepsilon_{jrs}[M^{lm}, M^{rs}] \\
&= \frac{i}{4}\varepsilon_{ilm}\varepsilon_{jrs}(M^{rm}g^{sl} + M^{ms}g^{lr} - M^{rl}g^{sm} - M^{ls}g^{mr}) \\
&= i\,\varepsilon_{ilm}\varepsilon_{jrs}M^{rm}g^{sl} = i\,\varepsilon_{ilm}\varepsilon_{jrl}M^{rm} \\
&= -i\,\varepsilon_{jrl}\varepsilon_{lim}M^{rm} = -i(\delta_{ji}\delta_{rm} - \delta_{jm}\delta_{ri})M^{rm} \\
&= iM^{ij}
\end{aligned}$$

where $M^{rm}\delta_{rm} = M^{rr} = 0$ since M^{rm} is antisymmetric. But we can invert the equation

$$J_i = \frac{1}{2}\varepsilon_{ijk}M^{jk}$$

to write

$$\varepsilon_{lmi}J_i = \frac{1}{2}\varepsilon_{lmi}\varepsilon_{ijk}M^{jk} = \frac{1}{2}(\delta_{lj}\delta_{mk} - \delta_{lk}\delta_{mj})M^{jk} = \frac{1}{2}(M^{lm} - M^{ml}) = M^{lm}$$

and therefore

$$[J_i, J_j] = i\varepsilon_{ijk}J_k. \tag{137a}$$

Similarly, we find

$$[J_i, K_j] = i\varepsilon_{ijk}K_k \tag{137b}$$

$$[K_i, K_j] = -i\varepsilon_{ijk}J_k \tag{137c}$$

It should be clear that $\mathbf{J}$ generates rotations and $\mathbf{K}$ generates boosts. It is also important to note that comparing equations (137b) and (126) we see that $\mathbf{K}$ is a polar (or true) vector, whereas $\mathbf{J}$ is an axial (or pseudo) vector. This is because letting $\mathbf{J} \to -\mathbf{J}$, the left side of (137a) does not change sign, whereas the right side does. This is also a consequence of the fact that $\mathbf{J}$ is essentially the angular momentum operator $\mathbf{r} \times \mathbf{p}$ which, because of the cross product, depends on the chosen orientation of the coordinate system for its definition. Indeed, under the parity operation $\mathbf{x} \to -\mathbf{x}$, the boost generators M^{i0} defined in Section 8 *will* change sign because we would have $\boldsymbol{\beta} \to -\boldsymbol{\beta}$, but the spatial rotation generators M^{ij} will *not* change sign. (This is essentially because under a rotation $d\boldsymbol{\theta}$ we have $d\mathbf{v} = d\boldsymbol{\theta} \times \mathbf{v}$ and this doesn't change sign if both $d\boldsymbol{\theta}$ and $\mathbf{v}$ change sign.)

Now let's write out the term

$$\omega_{\mu\nu} M^{\mu\nu} = 2\omega_{0i} M^{0i} + \omega_{ij} M^{ij} = -2\omega_{0i} K_i + \omega_{ij} \varepsilon_{ijk} J_k$$

so that we may write

$$U(\Lambda) = e^{\frac{i}{2}\omega_{\mu\nu} M^{\mu\nu}} = e^{-i\mathbf{a}\cdot\mathbf{J} + i\mathbf{b}\cdot\mathbf{K}}. \tag{138}$$

Define

$$\mathbf{A} = \frac{1}{2}(\mathbf{J} + i\mathbf{K}) \qquad \text{and} \qquad \mathbf{B} = \frac{1}{2}(\mathbf{J} - i\mathbf{K}) \tag{139}$$

or, equivalently,

$$\mathbf{J} = \mathbf{A} + \mathbf{B} \qquad \text{and} \qquad \mathbf{K} = -i(\mathbf{A} - \mathbf{B}). \tag{140}$$

The reason for this is that now $\mathbf{A}$ and $\mathbf{B}$ satisfy the simple commutation relations

$$[A_i, A_j] = i\varepsilon_{ijk} A_k \tag{141a}$$

$$[B_i, B_j] = i\varepsilon_{ijk} B_k \tag{141b}$$

$$[A_i, B_j] = 0 \tag{141c}$$

so they are commuting angular momentum generators. (These are just what we had in equations (95).)

Using equation (140) in (138) we have (by (141c) also)

$$U(\Lambda) = e^{\mathbf{A}\cdot(\mathbf{b} - i\mathbf{a})} e^{-\mathbf{B}\cdot(\mathbf{b} + i\mathbf{a})} \tag{142}$$

and we have (again) decomposed our Lorentz transformation into a product of two rotations, i.e., a direct product. It is also worth pointing out that the $U(\Lambda)$ of equation (142) is unitary if and only if $\mathbf{b} = 0$, i.e., only for a rotation.

Because of equations (141), the generators $\mathbf{A}$ and $\mathbf{B}$ are represented by $(2A+1)$- and $(2B+1)$-dimensional angular momentum matrices respectively. Treating $\mathbf{A}$ and $\mathbf{B}$ as independent variables, we then describe a particle of spin $\mathbf{J} = \mathbf{A} + \mathbf{B}$. We thus define the $[(2A+1) \times (2B+1)]$-dimensional irreducible representation (A, B) for any integer values of $2A$ and $2B$ by

$$\langle a'b'|\mathbf{A}|ab\rangle = \delta_{b'b} \mathbf{J}^{(A)}_{a'a} \qquad \text{and} \qquad \langle a'b'|\mathbf{B}|ab\rangle = \delta_{a'a} \mathbf{J}^{(B)}_{b'b} \tag{143}$$

where $a = -A, \ldots, A$, $b = -B, \ldots, B$ and the states are really $|ab\rangle = |a\rangle \otimes |b\rangle$. The operators are really $\mathbf{A} \otimes I$ and $I \otimes \mathbf{B}$ so that $\mathbf{J}^{(A)}_{a'a} = \langle a'|\mathbf{A}|a\rangle$ and similarly for $\mathbf{B}$. In other words, the operators $\mathbf{J}^{(j)}$ are the usual $(2j+1)$-dimensional representation of the rotation group, i.e.,

$$\langle j\sigma'|J_z|j\sigma\rangle = \delta_{\sigma'\sigma}\sigma$$

and

$$\langle j\sigma'|J_\pm|j\sigma\rangle = \sqrt{j(j+1) - \sigma(\sigma \pm 1)}\,\delta_{\sigma',\sigma\pm1}.$$

Although the representation (A, B) is in general reducible, as shown previously it is reducible for rotations alone. In other words, it is the product of $(2A+1)$- and $(2B+1)$-dimensional representations.

Now note that under space inversion (i.e., parity) we have $\mathbf{J} \to +\mathbf{J}$ (it's an axial vector), while $\mathbf{K} \to -\mathbf{K}$ (it's a polar vector). But then under space inversion we have $\mathbf{A} \to \mathbf{B}$ and $\mathbf{B} \to \mathbf{A}$ so the representation $(A, B) \to (B, A)$. Therefore, if we want to construct a wave function that has space-innversion symmetry, it must transform under a representation of the form (j, j) or, we must double the number of components and have them transform under $(j, 0) \oplus (0, j)$. The representations are often denoted by $D^{(A,B)}(\Lambda)$. The $(j, 0)$ representation corresponds to $\mathbf{A} = \mathbf{J}$ and $\mathbf{B} = 0$, and the $(2j+1)$-dimensional matrix representing a finite Lorentz transformation we be denoted by $D^{(j)}(\Lambda)$. Similarly, the $(0, j)$ representation corresponds to $\mathbf{A} = 0$ and $\mathbf{B} = \mathbf{J}$ and will be denoted by $\overline{D}^{(j)}(\Lambda)$. From equation (142) we see that the two representations are related by

$$D^{(j)}(\Lambda) = \overline{D}^{(j)}(\Lambda^{-1})^\dagger. \tag{144}$$

Let us now see how to describe the action of a Lorentz transformation on the states. Using equations (127) and (131) it is easy to verify that the operator

$$\mathscr{M}^2 = P_\mu P^\mu \tag{145}$$

commutes with all of the generators. (You will find it very helpful to use the general identity $[ab, c] = a[b, c] + [a, c]b$.) We may thus classify the irreducible representations of the group by the value of this invariant operator. Since the P^μ all commute, they may be simultaneously diagonalized, and hence we define the states $|p\rangle$ such that

$$P^\mu|p\rangle = p^\mu|p\rangle.$$

To interpret this, let's look at translations as they are described in the Schrödinger theory. If we consider a displacement $\mathbf{x} \to \mathbf{x}' = \mathbf{x} + \mathbf{a}$, then $\psi(\mathbf{x}) \to \psi'(\mathbf{x}') = \psi'(\mathbf{x} + \mathbf{a})$ or $\psi'(\mathbf{x}) = \psi(\mathbf{x} - \mathbf{a})$. Expanding the infinitesimal case with $\|\mathbf{a}\| \ll 1$ we have

$$\psi'(\mathbf{x}) = \psi(\mathbf{x} - \mathbf{a}) = \psi(\mathbf{x}) - \mathbf{a} \cdot \boldsymbol{\nabla}\psi(\mathbf{x}) = [1 - i\mathbf{a} \cdot \mathbf{P}]\psi(\mathbf{x})$$

where $\mathbf{P} = -i\boldsymbol{\nabla}$ is the momentum operator. For a finite displacement we exponentiate as usual to write

$$\psi'(\mathbf{x}) = e^{-i\mathbf{a}\cdot\mathbf{P}}\psi(\mathbf{x}).$$

Now, the effect of $U(a)$ (which generates translations by definition) acting on a state $|p\rangle$ is given by

$$U(a)|p\rangle = e^{ia_\mu P^\mu}|p\rangle = e^{ia_\mu p^\mu}|p\rangle = e^{ia\cdot p}|p\rangle$$

so that comparing this with the Schrödinger theory, we identify P^μ as the 4-momentum of the system. It is now clear that equation (145) represents the square of the system's mass.

Until further notice, all of what follows is based on the assumption that we are dealing with particles of nonzero rest mass.

From experience, we know that one-particle states are also characterized by their spin, i.e., the angular momentum in the rest frame of the particle. We now seek the invariant operator that describes spin. Such an operator is

$$W_\sigma := -\frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}M^{\mu\nu}P^\rho \tag{146}$$

which is called the **Pauli-Lubanski spin vector**. (We will discuss the representations of the Lorentz group in more detail below, and provide some motivation for this definition when we define the little group.) Note that for a system at rest we have $\mathbf{P} = 0$ and hence $W_\sigma = -(1/2)\varepsilon_{\mu\nu 0\sigma}M^{\mu\nu}P^0$. Therefore, for a system at rest we have $W_0 = 0$ and $W_i = -(1/2)\varepsilon_{\mu\nu 0 i}M^{\mu\nu}P^0 = -(1/2)\varepsilon_{0 i\mu\nu}M^{\mu\nu}P^0 = -(1/2)\varepsilon_{ijk}M^{jk}P^0$ so that from equation (132) we see that $W_i = -mJ_i$. We also clearly have

$$W_\sigma P^\sigma = 0 \tag{147}$$

and from equation (131) it also follows that

$$[W_\sigma, P^\mu] = 0. \tag{148}$$

Since W_σ is a 4-vector by construction, it must also have the same commutation relation with $M^{\mu\nu}$ that P_σ does (see the comment immediately following equation (131)), and hence

$$[M_{\mu\nu}, W_\sigma] = -i(W_\mu g_{\nu\sigma} - W_\nu g_{\mu\sigma}). \tag{149}$$

Then equations (148) and (149) may be used to easily prove

$$[W_\alpha, W_\beta] = i\varepsilon_{\alpha\beta\mu\nu}W^\mu P^\nu \tag{150}$$

and

$$[M_{\mu\nu}, W_\sigma W^\sigma] = 0. \tag{151}$$

Thus, like $\mathscr{M}^2 = P_\mu P^\mu$, equations (148) and (151) show that $\mathscr{S}^2 = W_\sigma W^\sigma$ commutes with all of the generators. These two invariants are called **Casimir operators**, and they are the only such operators for the Poincaré group. (It is a nontrivial theorem that the number of Casimir operators of a Lie group is equal to the rank of the group.)

We now have two operators which may be used to label the irreducible representations. To find the eigenvalues of $W_\sigma W^\sigma$ we go to the rest frame. Let

$L(p)$ be the Lorentz transformation which takes the 4-vector $(m, \mathbf{0})$ to p^μ. If $|\Psi_p\rangle$ is the state vector for a system with 4-momentum p^μ, then this system at rest is obtained by operating on $|\Psi_p\rangle$ with $U(L^{-1}(p))$, i.e.,

$$|\Psi_{\text{rest}}\rangle = U(L^{-1}(p))|\Psi_p\rangle.$$

Then we have

$$W_\sigma W^\sigma |\Psi_{\text{rest}}\rangle = W_\sigma W^\sigma U(L^{-1}(p))|\Psi_p\rangle = U(L^{-1}(p))W'_\sigma W'^\sigma |\Psi_p\rangle \tag{152}$$

where the operator in the rest frame is given by

$$W'_\sigma = U(L(p))W_\sigma U^{-1}(L(p))$$

and we used the fact that $U(L^{-1}(p)) = U^{-1}(L(p))$. Since W_σ is a 4-vector, we use equation (128) to conclude that

$$W'_\sigma = (L^{-1}(p))_\sigma{}^\rho W_\rho$$

or

$$W'^\sigma = (L^{-1}(p))^\sigma{}_\rho W^\rho.$$

It is also easy to show directly that $W'_\sigma W'^\sigma = W_\sigma W^\sigma$. Indeed, we have

$$W'_\sigma W'^\sigma = (L^{-1}(p))_\sigma{}^\alpha (L^{-1}(p))^\sigma{}_\beta W_\alpha W^\beta = L(p)^\alpha{}_\sigma (L^{-1})^\sigma{}_\beta W_\alpha W^\beta = W_\alpha W^\alpha$$

or alternatively

$$\begin{aligned} W'_\sigma W'^\sigma &= U(L(p))W_\sigma U^{-1}(L(p))U(L(p))W^\sigma U^{-1}(L(p)) \\ &= U(L(p))W_\sigma W^\sigma U^{-1}(L(p)) \\ &= W_\sigma W^\sigma \end{aligned}$$

because $\mathscr{S}^2 = W_\sigma W^\sigma$ commutes with all of the generators, and hence with any $U(a, \Lambda)$.

It is also worth noting that (since $U(L(p))$ is unitary)

$$\langle\Psi_{\text{rest}}|W_\sigma W^\sigma|\Psi_{\text{rest}}\rangle = \langle\Psi_p|U(L(p))W_\sigma W^\sigma U^{-1}(L(p))|\Psi_p\rangle = \langle\Psi_p|W_\sigma W^\sigma|\Psi_p\rangle$$

and therefore we have the equivalent expectation values

$$\begin{aligned} \langle\Psi_{\text{rest}}|W'_\sigma W'^\sigma|\Psi_{\text{rest}}\rangle = \langle\Psi_{\text{rest}}|W_\sigma W^\sigma|\Psi_{\text{rest}}\rangle &= \langle\Psi_p|W_\sigma W^\sigma|\Psi_p\rangle \\ &= \langle\Psi_p|W'_\sigma W'^\sigma|\Psi_p\rangle. \end{aligned}$$

The reason for pointing this out is that while they are all mathematically equivalent, we will show below that the operator $W'_\sigma W'^\sigma$ is just $-m^2\mathbf{S}^2$ which we think of as measuring the spin of the particle. Since spin is just the angular momentum in the rest frame, the comments following equation (123) suggest that from an intuitive viewpoint we want to consider the expectation value $\langle\Psi_{\text{rest}}|W_\sigma W^\sigma|\Psi_{\text{rest}}\rangle = \langle\Psi_p|W'_\sigma W'^\sigma|\Psi_p\rangle$.

Now recall that equation (51) gives the pure boost $\Lambda^\mu{}_\nu$ where $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ and x is the lab frame while x' is the moving (or rest) frame. In other words, this $\Lambda^\mu{}_\nu$ boosts *from* the frame with momentum p *to* the rest frame.

$$\Lambda^\mu{}_\nu = \begin{bmatrix} \gamma & -\gamma\beta^1 & -\gamma\beta^2 & -\gamma\beta^3 \\ -\gamma\beta^1 & 1+\frac{(\gamma-1)}{(\beta)^2}(\beta^1)^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^1\beta^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^1\beta^3 \\ -\gamma\beta^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^2\beta^1 & 1+\frac{(\gamma-1)}{(\beta)^2}(\beta^2)^2 & \frac{(\gamma-1)}{(\beta)^2}\beta^2\beta^3 \\ -\gamma\beta^3 & \frac{(\gamma-1)}{(\beta)^2}\beta^3\beta^1 & \frac{(\gamma-1)}{(\beta)^2}\beta^3\beta^2 & 1+\frac{(\gamma-1)}{(\beta)^2}(\beta^3)^2 \end{bmatrix}.$$

To boost from the rest frame to the lab frame we have $x^\mu = (\Lambda^{-1})^\mu{}_\nu x'^\nu$ where $(\Lambda^{-1})^\mu{}_\nu = (\Lambda^T)^\mu{}_\nu = \Lambda_\nu{}^\mu := (L(p))^\mu{}_\nu$. Then $(L^{-1}(p))^\mu{}_\nu = (L(p))_\nu{}^\mu = \Lambda^\mu{}_\nu$.

Now we use the relativistic expressions $E_{\mathbf{p}} = \gamma m$, $p^i = \gamma m\beta^i$ and $\mathbf{p}^2 = E_{\mathbf{p}}^2 - m^2 = (E_{\mathbf{p}} - m)(E_{\mathbf{p}} + m)$ to write

$$\gamma = \frac{E_{\mathbf{p}}}{m} \qquad \text{and} \qquad -\gamma\beta^i = -\frac{p^i}{m} \tag{153}$$

along with (since $(\beta)^2 = \mathbf{p}^2/\gamma^2 m^2$)

$$\frac{(\gamma-1)}{(\beta)^2}\beta^i\beta^j = \frac{(E_{\mathbf{p}}/m - 1)}{(\mathbf{p}^2/\gamma^2 m^2)}\frac{p^i p^j}{\gamma^2 m^2} = \frac{E_{\mathbf{p}} - m}{m\mathbf{p}^2}p^i p^j = \frac{p^i p^j}{m(E_{\mathbf{p}} + m)}.$$

Then in block matrix form we have

$$(L^{-1}(p))^\mu{}_\nu = \left[\begin{array}{c|c} E_{\mathbf{p}}/m & -p^i/m \\ \hline -p^i/m & g^i_j + \frac{p^i p^j}{m(E_{\mathbf{p}}+m)} \end{array}\right].$$

We can now write out the components of W'^σ as follows (note $p^j = -p_j$ and $P_\sigma W^\sigma = E_{\mathbf{p}} W^0 - \mathbf{p}\cdot\mathbf{W} = 0$):

$$W'^0 = (L^{-1}(p))^0{}_\rho W^\rho = (E_{\mathbf{p}}/m)W^0 - \mathbf{p}\cdot\mathbf{W}/m = (1/m)P_\sigma W^\sigma = 0$$

$$\begin{aligned} W'^i &= (L^{-1}(p))^i{}_\rho W^\rho = (L^{-1}(p))^1{}_0 W^0 + (L^{-1}(p))^i{}_j W^j \\ &= -\frac{p^i}{m}W^0 + W^i - \frac{p^i p_j W^j}{m(E_{\mathbf{p}}+m)} = -\frac{p^i}{m}W^0 + W^i + \frac{p^i(\mathbf{p}\cdot\mathbf{W})}{m(E_{\mathbf{p}}+m)} \\ &= W^i - \frac{p^i}{m}\left[W^0 - \frac{\mathbf{p}\cdot\mathbf{W}}{E_{\mathbf{p}}+m}\right] = W^i - \frac{p^i W^0}{m}\left[1 - \frac{E_{\mathbf{p}}}{E_{\mathbf{p}}+m}\right] \\ &= W^i - \frac{p^i}{E_{\mathbf{p}}+m}W^0. \end{aligned}$$

Since spin is the angular momentum in the rest frame, let us now define the operators

$$S_i = -\frac{1}{m}W'^i. \tag{154}$$

It can be shown, after a lot of algebra (see below for the details), that

$$[S_i, S_j] = i\varepsilon_{ijk}S_k \tag{155}$$

which is just the usual commutation relation for an angular momentum operator $\mathbf{S}$ where the eigenvalues of $\mathbf{S}^2$ are equal to $s(s+1)$ with $s = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$. Since $W'^0 = 0$, we now have

$$W'_\sigma W'^\sigma = W'_i W'^i = -m^2\mathbf{S}^2$$

so that (see equation (152))

$$W_\sigma W^\sigma|\Psi_{\text{rest}}\rangle = U(L^{-1}(p))W'_\sigma W'^\sigma|\Psi_p\rangle = -m^2 s(s+1)U(L^{-1}(p))|\Psi_p\rangle$$

or

$$W_\sigma W^\sigma|\Psi_{\text{rest}}\rangle = -m^2 s(s+1)|\Psi_{\text{rest}}\rangle. \tag{156}$$

In summary, the irreducible representations of the Poincaré group are characterized by two invariants: mass and spin.

Now let's go back and prove equation (155). This is equivalent to showing that

$$[W'_i, W'_j] = -im\varepsilon_{ijk}W'^k$$

because from the definition of S_i we have $S_i = (-1/m)W'^i = (+1/m)W'_i$ so that

$$[W'_i, W'_j] = m^2[S_i, S_j] = m^2 i\varepsilon_{ijk}S_k = -im\varepsilon_{ijk}W'^k.$$

We start from

$$W'_i = W_i - \frac{p_i}{E+m}W_0$$

where now p_i is a c-number since our states are eigenstates of P_μ, and where for simplicity we write E instead of $E_{\mathbf{p}}$. Then using equations (146), (149) and (147) we have (note that now $\varepsilon_{\mu\nu\rho\sigma}$ carries spacetime (tensor) indices because of equation (146) where both sides are tensors)

$$\begin{aligned}
[W'_i, W'_j] &= [W_i - \frac{p_i}{E+m}W_0, W_j - \frac{p_j}{E+m}W_0] \\
&= [W_i, W_j] - \frac{p_j}{E+m}[W_i, W_0] - \frac{p_i}{E+m}[W_0, W_j] \\
&= -\frac{1}{2}\varepsilon^{\mu\nu}{}_{\rho i}p^\rho[M_{\mu\nu}, W_j] + \frac{1}{2}\frac{p_j p^\rho}{E+m}\varepsilon^{\mu\nu}{}_{\rho i}[M_{\mu\nu}, W_0] \\
&\qquad\qquad - \frac{1}{2}\frac{p_i p^\rho}{E+m}\varepsilon^{\mu\nu}{}_{\rho j}[M_{\mu\nu}, W_0] \\
&= \frac{i}{2}\varepsilon^{\mu\nu}{}_{\rho i}p^\rho(W_\mu g_{\nu j} - W_\nu g_{\mu j}) - \frac{i}{2}\frac{p_j p^\rho}{E+m}\varepsilon^{\mu\nu}{}_{\rho i}(W_\mu g_{\nu 0} - W_\nu g_{\mu 0}) \\
&\qquad\qquad + \frac{i}{2}\frac{p_i p^\rho}{E+m}\varepsilon^{\mu\nu}{}_{\rho j}(W_\mu g_{\nu 0} - W_\nu g_{\mu 0})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{i}{2}(\varepsilon^{\mu}{}_{j\rho i}p^{\rho}W_{\mu} - \varepsilon_{j}{}^{\nu}{}_{\rho i}p^{\rho}W_{\nu}) \\
&\qquad - \frac{i}{2}\frac{p^{\rho}}{E+m}(p_j\varepsilon^{\mu}{}_{0\rho i}W_{\mu} - p_j\varepsilon_{0}{}^{\nu}{}_{\rho i}W_{\nu} - p_i\varepsilon^{\mu}{}_{0\rho j}W_{\mu} + p_i\varepsilon_{0}{}^{\nu}{}_{\rho j}W_{\nu}) \\
&= i\varepsilon^{\mu}{}_{j\rho i}p^{\rho}W_{\mu} - i\frac{p^{\rho}}{E+m}(p_j\varepsilon^{\mu}{}_{0\rho i}W_{\mu} - p_i\varepsilon^{\mu}{}_{0\rho j}W_{\mu}) \\
&= ip^{\rho}W^{\mu}\Big(\varepsilon_{\mu j\rho i} - \frac{p_j}{E+m}\varepsilon_{\mu 0\rho i} + \frac{p_i}{E+m}\varepsilon_{\mu 0\rho j}\Big).
\end{aligned}
$$

Note that this last line vanishes identically if $i = j$, so we now assume that $i \neq j$ and expand the sums over μ and ρ, keeping in mind that the ε symbols restrict the possible values of the indices. For example, since $i \neq j$ we have $p^{\rho}W^{\mu}\varepsilon_{\mu j\rho i} = p^{\rho}W^{0}\varepsilon_{0j\rho i} + p^{\rho}W^{k}\varepsilon_{kj\rho i}$ where there is no sum over k because it must be either 1, 2 or 3 depending on what i and j are. But now summing over ρ we really don't have any choices, so that $p^{\rho}W^{\mu}\varepsilon_{\mu j\rho i} = p^{k}W^{0}\varepsilon_{0jki} + p^{0}W^{k}\varepsilon_{kj0i}$ where we stress that there is no sum over the repeated index k.

Continuing with this process we have (using the fact that $p^0 = E$ and remembering that there is no sum over any of the indices i, j, k)

$$
\begin{aligned}
[W'_i, W'_j] &= ip^{\rho}W^{0}\varepsilon_{0j\rho i} + ip^{\rho}W^{k}\varepsilon_{kj\rho i} - i\frac{p^{\rho}p_j}{E+m}(W^{j}\varepsilon_{j0\rho i} + W^{k}\varepsilon_{k0\rho i}) \\
&\qquad + i\frac{p^{\rho}p_i}{E+m}(W^{i}\varepsilon_{i0\rho j} + W^{k}\varepsilon_{k0\rho j}) \\
&= i\varepsilon_{0jki}p^{k}W^{0} + iEW^{k}\varepsilon_{kj0i} - i\frac{p_j}{E+m}(W^{j}\varepsilon_{j0ki}p^{k} + W^{k}\varepsilon_{k0ji}p^{j}) \\
&\qquad + i\frac{p_i}{E+m}(W^{i}\varepsilon_{i0kj}p^{k} + W^{k}\varepsilon_{k0ij}p^{i}).
\end{aligned}
$$

Now use the antisymmetry of ε and note that ε_{0ijk} is exactly the same as ε_{ijk} to write

$$
\begin{aligned}
[W'_i, W'_j] = i\varepsilon_{ijk}\Big[&W^{0}p^{k} - W^{k}E \\
&+ \Big(\frac{1}{E+m}\Big)(W^{j}p_jp^{k} - W^{k}p^{j}p_j + W^{i}p_ip^{k} - W^{k}p^{i}p_i)\Big].
\end{aligned}
$$

Next look at the four terms in parenthesis. By adding and subtracting the additional term $W^{k}p_kp^{k}$ we can rewrite this term as

$$
\begin{aligned}
(W^{i}p_i + W^{j}p_j + W^{k}p_k)p^{k} &- W^{k}(p^{i}p_i + p^{j}p_j + p^{k}p_k) \\
&= (W^{\mu}p_{\mu} - W^{0}E)p^{k} + W^{k}\mathbf{p}^2
\end{aligned}
$$

where we also used $\sum_{l=1}^{3} p^{l}p_l = -\sum p^{l}p^{l} = -\mathbf{p}^2$. But $W^{\mu}p_{\mu} = 0$, and recalling that $E^2 = \mathbf{p}^2 + m^2$ we finally have

$$
[W'_i, W'_j] = i\varepsilon_{ijk}\Big[W^{0}p^{k} - W^{k}E + \Big(\frac{1}{E+m}\Big)(-W^{0}Ep^{k} + W^{k}\mathbf{p}^2)\Big]
$$

$$
\begin{aligned}
&= i\varepsilon_{ijk}\Big[W^0p^k - W^k\Big(E - \frac{\mathbf{p}^2}{E+m}\Big) - W^0p^k\Big(\frac{E}{E+m}\Big)\Big] \\
&= i\varepsilon_{ijk}\Big[W^0p^k - W^k\Big(\frac{E^2+mE-\mathbf{p}^2}{E+m}\Big) - W^0p^k\Big(\frac{E}{E+m}\Big)\Big] \\
&= i\varepsilon_{ijk}\Big[W^0p^k - mW^k - W^0p^k\Big(\frac{E}{E+m}\Big)\Big] \\
&= i\varepsilon_{ijk}\Big[W^0p^k\Big(1 - \frac{E}{E+m}\Big) - mW^k\Big] \\
&= i\varepsilon_{ijk}m\Big[\frac{W^0p^k}{E+m} - W^k\Big] \\
&= -im\varepsilon_{ijk}W'^k
\end{aligned}
$$

which is what we wanted to show.

We still have to motivate the definition (146) of the Pauli-Lubanski spin vector. This is based on what is called the "method of induced representations" of the inhomogeneous Lorentz group.

Since all of the P^μ commute, we can describe our states as eigenstates of the 4-momentum. Let us label these states as $|\Psi_{p,\sigma}\rangle$ where σ stands for all other degrees of freedom necessary to describe the state. Then by definition we have

$$P^\mu|\Psi_{p,\sigma}\rangle = p^\mu|\Psi_{p,\sigma}\rangle.$$

From equation (122a) we see how these states transform under translations:

$$U(a)|\Psi_{p,\sigma}\rangle = e^{ia_\mu p^\mu}|\Psi_{p,\sigma}\rangle.$$

What we now need to figure out is how they transform under homogeneous Lorentz transformations $U(\Lambda)$.

Using equation (129) we see that

$$
\begin{aligned}
P^\mu[U(\Lambda)|\Psi_{p,\sigma}\rangle] &= U(\Lambda)[U^{-1}(\Lambda)P^\mu U(\Lambda)]|\Psi_{p,\sigma}\rangle = U(\Lambda)[\Lambda^\mu{}_\nu P^\nu]|\Psi_{p,\sigma}\rangle \\
&= \Lambda^\mu{}_\nu p^\nu[U(\Lambda)|\Psi_{p,\sigma}\rangle]
\end{aligned}
$$

so that the state $U(\Lambda)|\Psi_{p,\sigma}\rangle$ is an eigenstate of P^μ with eigenvalue $\Lambda^\mu{}_\nu p^\nu$ and hence must be some linear combination of the states $|\Psi_{\Lambda p,\sigma}\rangle$. Thus we write

$$U(\Lambda)|\Psi_{p,\sigma}\rangle = \sum_{\sigma'} C_{\sigma'\sigma}(\Lambda,p)|\Psi_{\Lambda p,\sigma'}\rangle. \tag{157}$$

We think of the $|\Psi_{p,\sigma}\rangle$ as forming a basis for the space carrying a representation of the Poincaré group, and therefore in general the matrix $C_{\sigma'\sigma}(\Lambda,p)$ will be complicated. However, recall from basic linear algebra that we can diagonalize a matrix by finding suitable linear combinations of the basis vectors (these will then form a basis of eigenvectors) relative to which the matrix is diagonal. But if any eigenvalue has multiplicity greater than one, then all we can do is put the matrix into block diagonal form where each block is an invariant subspace. In our present situation, we will assume that by using suitable linear combinations of the $C_{\sigma'\sigma}(\Lambda, p)$ it is possible to choose the labels σ so that the matrix $C_{\sigma'\sigma}(\Lambda, p)$ will be block diagonal, and hence for σ in the appropriate range, each block will form an irreducible representation of the Poincaré group. We will identify the states of a specific type of particle with each of these irreducible representations.

To analyze the structure of the coefficients $C_{\sigma'\sigma}(\Lambda, p)$ in a particular irreducible representation, first note that all proper orthochronous Lorentz transformations leave both the quantity $p_\mu p^\mu$ and the sign of p^0 invariant. Hence for each value of p^2 and each sign of p^0, we can choose a "standard" 4-momentum $\mathring{p}^\mu$ and write any other p^μ in this same class as

$$p^\mu = L^\mu{}_\nu(p)\mathring{p}^\nu \tag{158}$$

where $L^\mu{}_\nu(p)$ is some standard Lorentz transformation that depends both on p^μ and our reference $\mathring{p}^\mu$. (The six classes are $p^2 > 0$ and $p^2 = 0$ together with each sign of p^0, along with $p^2 < 0$ and $p^\mu = 0$. But only the three classes $p^2 > 0$ with $p^0 > 0$, $p^2 = 0$ with $p^0 > 0$, and $p^\mu = 0$ have any known physical interpretation.) We then *define* the states $|\Psi_{p,\sigma}\rangle$ by

$$|\Psi_{p,\sigma}\rangle := N(p)U(L(p))|\Psi_{\mathring{p},\sigma}\rangle \tag{159}$$

where $N(p)$ is a normalization factor to be determined.

Observe that $p = L(p)\mathring{p}$ so that $\Lambda p = L(\Lambda p)\mathring{p}$ where Λ is any homogeneous Lorentz transformation. Then $\mathring{p} = L^{-1}(\Lambda p)\Lambda p = L^{-1}(\Lambda p)\Lambda L(p)\mathring{p}$ so that the transformation $L^{-1}(\Lambda p)\Lambda L(p)$ takes $\mathring{p}$ back to $\mathring{p}$. This transformation then belongs to the subgroup of the homogeneous Lorentz group consisting of those transformations $\mathring{\Lambda}^\mu{}_\nu$ that leave $\mathring{p}$ invariant:

$$\mathring{\Lambda}^\mu{}_\nu \mathring{p}^\nu = \mathring{p}^\mu. \tag{160}$$

This subgroup is called the **little group**.

For example, if we consider the class defined by $p^2 > 0$ and $p^0 > 0$, then we can take $\mathring{p}^\mu = (m, 0, 0, 0)$ so that $\mathring{p}^2 = m^2 > 0$ and $\mathring{p}^0 = m > 0$. Then the little group is just $SO(3)$ (the ordinary group of rotations in $\mathbb{R}^3$ with positive determinant) because rotations are the only proper orthochronous Lorentz transformations that leave at rest a particle with $\mathbf{p} = 0$.

For an infinitesimal transformation $\mathring{\Lambda}^\mu{}_\nu = g^\mu{}_\nu + \mathring{\omega}^\mu{}_\nu$ we then must have

$$\mathring{\omega}^\mu{}_\nu \mathring{p}^\nu = 0 \tag{161}$$

and a general expression for $\mathring{\omega}_{\mu\nu}$ that satisfies this is

$$\mathring{\omega}_{\mu\nu} = \varepsilon_{\mu\nu\rho\sigma}\mathring{p}^\rho n^\sigma$$

where n^σ is an arbitrary 4-vector. The corresponding unitary transformation $U(\mathring{\Lambda})$ is then

$$U(\mathring{\Lambda}) = 1 + \frac{i}{2}\mathring{\omega}_{\mu\nu}M^{\mu\nu} = 1 + \frac{i}{2}\varepsilon_{\mu\nu\rho\sigma}\mathring{p}^\rho n^\sigma M^{\mu\nu}$$

or

$$U(\mathring{\Lambda}) = 1 - in^\sigma W_\sigma$$

where

$$W_\sigma = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}M^{\mu\nu}P^\rho$$

is the Pauli-Lubanski spin vector. Note that we replaced $\mathring{p}^\rho$ by P^ρ because $U(\mathring{\Lambda})$ acts only on the states $|\Psi_{\mathring{p},\sigma}\rangle$. This is the motivation for equation (146) that we were looking for.

Since $\mathring{\Lambda}\mathring{p} = \mathring{p}$, we see from equation (157) that we may write

$$U(\mathring{\Lambda})|\Psi_{\mathring{p},\sigma}\rangle = \sum_{\sigma'} D_{\sigma'\sigma}(\mathring{\Lambda})|\Psi_{\mathring{p},\sigma'}\rangle \tag{162}$$

where the $D_{\sigma'\sigma}(\mathring{\Lambda})$ form a representation of the little group. Indeed, for any $\mathring{\Lambda}, \mathring{\Lambda}'$ we have

$$\begin{aligned}\sum_{\sigma'} D_{\sigma'\sigma}(\mathring{\Lambda}'\mathring{\Lambda})|\Psi_{\mathring{p},\sigma'}\rangle &= U(\mathring{\Lambda}'\mathring{\Lambda})|\Psi_{\mathring{p},\sigma}\rangle = U(\mathring{\Lambda}')U(\mathring{\Lambda})|\Psi_{\mathring{p},\sigma}\rangle \\ &= U(\mathring{\Lambda}')\sum_{\sigma''} D_{\sigma''\sigma}(\mathring{\Lambda})|\Psi_{\mathring{p},\sigma''}\rangle \\ &= \sum_{\sigma'\sigma''} D_{\sigma''\sigma}(\mathring{\Lambda})D_{\sigma'\sigma''}(\mathring{\Lambda}')|\Psi_{\mathring{p},\sigma'}\rangle\end{aligned}$$

which shows that

$$D_{\sigma'\sigma}(\mathring{\Lambda}'\mathring{\Lambda}) = \sum_{\sigma''} D_{\sigma'\sigma''}(\mathring{\Lambda}')D_{\sigma''\sigma}(\mathring{\Lambda})$$

and hence that the $D_{\sigma'\sigma}(\mathring{\Lambda})$ are in fact a representation of the little group (i.e., a homomorphism from the group of all such $\mathring{\Lambda}$ to the group of matrices $D_{\sigma'\sigma}(\mathring{\Lambda})$).

Acting on equation (159) with an arbitrary $U(\Lambda)$ we have (by inserting the identity transformation appropriately)

$$\begin{aligned}U(\Lambda)|\Psi_{p,\sigma}\rangle &= N(p)U(\Lambda)U(L(p))|\Psi_{\mathring{p},\sigma}\rangle = N(p)U(\Lambda L(p))|\Psi_{\mathring{p},\sigma}\rangle \\ &= N(p)U(L(\Lambda p))U(L^{-1}(\Lambda p)\Lambda L(p))|\Psi_{\mathring{p},\sigma}\rangle.\end{aligned}$$

Let us define the little group transformation known as the **Wigner rotation**

$$W(\Lambda, p) := L^{-1}(\Lambda p)\Lambda L(p)$$

(which was the original $\mathring{\Lambda}$ used as an example to define the little group) so using equation (162) this last equation becomes

$$\begin{aligned}U(\Lambda)|\Psi_{p,\sigma}\rangle &= N(p)U(L(\Lambda p))U(W(\Lambda,p))|\Psi_{\mathring{p},\sigma}\rangle \\ &= N(p)\sum_{\sigma'} D_{\sigma'\sigma}(W(\Lambda,p))U(L(\Lambda p))|\Psi_{\mathring{p},\sigma'}\rangle\end{aligned}$$

or, using equations (159) and (158)

$$U(\Lambda)|\Psi_{p,\sigma}\rangle = \frac{N(p)}{N(\Lambda p)}\sum_{\sigma'} D_{\sigma'\sigma}(W(\Lambda,p))|\Psi_{\Lambda p,\sigma'}\rangle. \tag{163}$$

Thus we see that, aside from determining the normalization factors, we have reduced the problem of finding the coefficients $C_{\sigma'\sigma}$ in equation (157) to the problem of finding the representations of the little group. This is the **method of induced representations**. For our present purposes, this is as far as we really need to go with this.

However, we can take a look at the normalization because it will give us two important basic results that are of great use in quantum field theory. We first want to show that when integrating over $\mathbf{p}$ "on the mass shell" (meaning that $p^2 = m^2$), the invariant volume element is

$$\frac{d^3\mathbf{p}}{(2\pi)^3 2\omega_{\mathbf{p}}}$$

where $\omega_{\mathbf{p}} = +\sqrt{\mathbf{p}^2 + m^2}$. (The numerical factors aren't necessary for Lorentz invariance.) To see this, first observe that

$$\frac{d^4p}{(2\pi)^4}2\pi\delta(p^2 - m^2)\theta(p_0)$$

is manifestly Lorentz invariant for proper orthochronous Lorentz transformations. (The **step function** is defined by $\theta(x) = 1$ for $x \geq 0$ and $\theta(x) = 0$ for $x < 0$.) This is because with $\det\Lambda = 1$ we have $d^4p' = |\partial p'/\partial p|\, d^4p = d^4p$, and for timelike p^μ an orthochronous transformation can't change the sign of p^0 so the step function doesn't change. Then we have

$$\begin{aligned}\frac{d^4p}{(2\pi)^4}2\pi\delta(p^2 - m^2)\theta(p_0) &= \frac{d^4p}{(2\pi)^3}\delta(p_0^2 - \omega_{\mathbf{p}}^2)\theta(p_0) \\ &= \frac{d^3\mathbf{p}\,dp_0}{(2\pi)^3}\delta[(p_0 - \omega_{\mathbf{p}})(p_0 + \omega_{\mathbf{p}})]\theta(p_0) \\ &= \frac{d^3\mathbf{p}\,dp_0}{(2\pi)^3}\,\frac{1}{2\omega_{\mathbf{p}}}\,[\delta(p_0 - \omega_{\mathbf{p}}) + \delta(p_0 + \omega_{\mathbf{p}})]\theta(p_0) \\ &= \frac{d^3\mathbf{p}\,dp_0}{(2\pi)^3 2\omega_{\mathbf{p}}}\delta(p_0 - \omega_{\mathbf{p}}) \\ &= \frac{d^3\mathbf{p}}{(2\pi)^3 2\omega_{\mathbf{p}}}\end{aligned}$$

where the first line is because $p^2 - m^2 = p_0^2 - \mathbf{p}^2 - m^2 = p_0^2 - \omega_{\mathbf{p}}^2$; the third line follows because integrating over all p_0 gives contributions at both $p_0 = +\omega_{\mathbf{p}}$ and $p_0 = -\omega_{\mathbf{p}}$, and using $\delta(ax) = (1/|a|)\delta(x)$; the fourth line is because the theta function restricts the range of integration to $p_0 > 0$ so $\delta(p_0 + \omega_{\mathbf{p}})$ always vanishes; and the last line follows because we want to integrate a function over d^3p so integrating over p_0 with the delta function just gives 1. This result is really a shorthand notation for

$$\frac{d^3\mathbf{p}}{(2\pi)^3 2\omega_{\mathbf{p}}} = \int_{p_0} \frac{d^4p}{(2\pi)^4} 2\pi\delta(p^2 - m^2)\theta(p_0). \tag{164}$$

Another way to see this is to write the integral of an arbitrary function $f(p) = f(p_0, \mathbf{p})$ as

$$\begin{aligned}\int d^4p\,\delta(p^2 - m^2)\theta(p_0)f(p) &= \int d^3\mathbf{p}\,dp_0\,\delta(p_0^2 - \mathbf{p}^2 - m^2)\theta(p_0)f(p_0, \mathbf{p}) \\ &= \int d^3\mathbf{p}\,\frac{1}{2\omega_{\mathbf{p}}}\delta(p_0 - \omega_p)f(p_0, \mathbf{p}) \\ &= \int \frac{d^3\mathbf{p}}{2\omega_{\mathbf{p}}} f(\omega_p, \mathbf{p})\end{aligned}$$

which also shows that when integrating over the mass shell $p^2 = m^2$ the invariant volume element is $d^3\mathbf{p}/\omega_{\mathbf{p}} = d^3\mathbf{p}/\sqrt{\mathbf{p}^2 + m^2}$.

The second result we want to show is that the **invariant delta function** $(2\pi)^3 2\omega_{\mathbf{p}}\delta(\mathbf{p} - \mathbf{q})$ is Lorentz invariant (where $p^2 = q^2 = m^2$ and again the numerical factors are not necessary). The easy way to see this is to use the definition of the delta function to write

$$f(\mathbf{p}) = \int d^3\mathbf{q}\,\delta(\mathbf{p} - \mathbf{q})f(\mathbf{q}) = \int \frac{d^3\mathbf{q}}{2\omega_{\mathbf{q}}}[2\omega_{\mathbf{q}}\delta(\mathbf{p} - \mathbf{q})]f(\mathbf{q}).$$

But $d^3\mathbf{q}/2\omega_{\mathbf{q}}$ is Lorentz invariant, so $2\omega_{\mathbf{q}}\delta^3(\mathbf{p} - \mathbf{q})$ must also be invariant.

The hard way to see this is a brute force calculation which is a good illustration of how to manipulate the delta function. Consider a boost in the p_3 direction:

$$p_1' = p_1 \qquad p_2' = p_2 \qquad p_3' = \gamma(p_3 + \beta p_0)$$

i.e., $\mathbf{p}_\perp' = \mathbf{p}_\perp$ and similarly for $\mathbf{q}$. (These are the usual boost equations if you remember that $p_0 = p^0$ while $p_i = -p^i$.) Now write

$$\begin{aligned}\delta(\mathbf{p} - \mathbf{q}) &= \delta(\mathbf{p}_\perp - \mathbf{q}_\perp)\delta(p_3 - q_3) = \delta(\mathbf{p}_\perp' - \mathbf{q}_\perp')\delta(p_3(p_3') - q_3) \\ &:= \delta(\mathbf{p}_\perp' - \mathbf{q}_\perp')\delta(f(p_3')).\end{aligned}$$

If $p_{3,0}'$ is a zero of $f(p_3')$, then expanding to first order we have

$$f(p_3') = f(p_{3,0}') + \frac{df}{dp_3'}(p_{3,0}')(p_3' - p_{3,0}') = \frac{df}{dp_3'}(p_{3,0}')(p_3' - p_{3,0}')$$

and therefore

$$\delta(f(p_3')) = \frac{1}{\left|\frac{df}{dp_3'}\right|_{p_{3,0}'}} \delta(p_3' - p_{3,0}').$$

To find the zero of $f(p_3')$ we first expand

$$\begin{aligned}0 = f(p_3') &= \gamma(p_3' - \beta p_0') - q_3 = \gamma\Big(p_3' - \beta\sqrt{\mathbf{p}'^2 + m^2}\Big) - q_3 \\ &= \gamma\Big(p_3' - \beta\sqrt{\mathbf{p}_\perp'^{\,2} + p_3'^{\,2} + m^2}\Big) - q_3.\end{aligned}$$

To solve this for p_3' we rearrange it to write

$$q_3 - \gamma p_3' = -\gamma\beta\sqrt{\mathbf{p}_\perp'^{\,2} + p_3'^{\,2} + m^2}$$

and then square both sides:

$$q_3^2 + \gamma^2 p_3'^{\,2} - 2\gamma q_3 p_3' = \gamma^2\beta^2\big(\mathbf{p}_\perp'^{\,2} + p_3'^{\,2} + m^2\big).$$

Grouping the $p_3'^{\,2}$ terms and using $\gamma^2(1-\beta^2) = 1$ we have

$$p_3'^{\,2} - 2\gamma q_3 p_3' + q_3{}^2 - \gamma^2\beta^2\big(\mathbf{p}_\perp'^{\,2} + m^2\big) = 0$$

which is just a quadratic in p_3'. Using the quadratic formula we obtain

$$\begin{aligned}p_{3,0}' &= \gamma q_3 \pm \sqrt{\gamma^2 q_3{}^2 - q_3{}^2 + \gamma^2\beta^2\big(\mathbf{p}_\perp'^{\,2} + m^2\big)} \\ &= \gamma q_3 \pm \gamma\beta\sqrt{q_3{}^2 + \mathbf{p}_\perp'^{\,2} + m^2}\end{aligned}$$

where we used $\gamma^2 - 1 = \gamma^2\beta^2$.

Next we note that the factor $\delta(\mathbf{p}_\perp - \mathbf{q}_\perp)$ in the expression for $\delta(\mathbf{p}-\mathbf{q})$ means that we must have $\mathbf{q}_\perp = \mathbf{p}_\perp = \mathbf{p}_\perp'$ so that (since $p^2 = q^2 = m^2$)

$$p_{3,0}' = \gamma q_3 \pm \gamma\beta q_0$$

so we take $p_{3,0}' = \gamma(q_3 + \beta q_0) \equiv q_3'$. Now we evaluate dp_3/dp_3' using $p_0' = \sqrt{\mathbf{p}'^2 + m^2}$:

$$\begin{aligned}\frac{dp_3}{dp_3'} &= \frac{d}{dp_3'}\gamma(p_3' - \beta p_0') = \gamma - \gamma\beta\frac{dp_0'}{dp_3'} = \gamma - \gamma\beta\frac{p_3'}{p_0'} \\ &= \frac{\gamma}{p_0'}(p_0' - \beta p_3') = \frac{p_0}{p_0'} = \frac{\omega_{\mathbf{p}}}{\omega_{\mathbf{p}'}}.\end{aligned}$$

Finally, we put all of this together to find

$$\begin{aligned}\delta(\mathbf{p}-\mathbf{q}) &= \delta(\mathbf{p}_\perp - \mathbf{q}_\perp)\delta(p_3 - q_3) \\ &= \delta(\mathbf{p}_\perp' - \mathbf{q}_\perp')\frac{1}{\left|\frac{df}{dp_3'}\right|_{p_3'=q_3'}}\delta(p_3' - q_3') = \frac{1}{\left|\frac{df}{dp_3'}\right|_{p_3'=q_3'}}\delta(\mathbf{p}' - \mathbf{q}') \\ &= \frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}}\delta(\mathbf{p}' - \mathbf{q}')\end{aligned}$$

and hence

$$\omega_{\mathbf{p}}\delta(\mathbf{p}-\mathbf{q}) = \omega_{\mathbf{p}'}\delta(\mathbf{p}'-\mathbf{q}') \tag{165}$$

as claimed.

Let us now return to determining the normalization constants in equation (163). To begin with, we could choose a different standard state $\mathring{p}'^{\mu}$ from the original one $\mathring{p}^{\mu}$ we chose (but still with $\mathring{p}'^2 = \mathring{p}^2$). However, we choose our standard states to have the normalization

$$\langle \Psi_{\mathring{p}',\sigma'} | \Psi_{\mathring{p},\sigma} \rangle = \delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}})\delta_{\sigma'\sigma}.$$

For arbitrary momentum p^{μ} and p'^{μ} (but still in the same irreducible representation with $p'^2 = p^2 = \mathring{p}^2$) we then have (using equation (159) and the fact that U is unitary)

$$\begin{aligned}
\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle &= \langle \Psi_{p',\sigma'} | N(p) U(L(p)) | \Psi_{\mathring{p},\sigma} \rangle \\
&= N(p) \langle U^{\dagger}(L(p)) \Psi_{p',\sigma} | \Psi_{\mathring{p},\sigma} \rangle \\
&= N(p) \langle U(L^{-1}(p)) \Psi_{p',\sigma} | \Psi_{\mathring{p},\sigma} \rangle
\end{aligned} \tag{166}$$

From equation (163) we see that

$$U(L^{-1}(p)) | \Psi_{p',\sigma'} \rangle = \frac{N(p')}{N(\mathring{p}')} \sum_{\sigma''} D_{\sigma''\sigma'}(W(L^{-1}(p), p')) | \Psi_{\mathring{p}',\sigma''} \rangle$$

where we have *defined* $\mathring{p}' := L^{-1}(p)p'$. Using the adjoint of this equation in equation (166) we have

$$\begin{aligned}
\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle &= \frac{N(p)N^*(p')}{N^*(\mathring{p}')} \sum_{\sigma''} D^*_{\sigma''\sigma'}(W(L^{-1}(p), p')) \langle \Psi_{\mathring{p}',\sigma''} | \Psi_{\mathring{p},\sigma} \rangle \\
&= \frac{N(p)N^*(p')}{N^*(\mathring{p}')} \sum_{\sigma''} D^*_{\sigma''\sigma'}(W(L^{-1}(p), p')) \delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}}) \delta_{\sigma''\sigma}.
\end{aligned}$$

Because of the factor $\delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}})$, the right side of this equation is nonzero only for $\mathring{\mathbf{p}}' = \mathring{\mathbf{p}}$, so we can take the constant $N^*(\mathring{p}')$ to be the same as $N^*(\mathring{p})$. But from equation (159) it is clear that $N(\mathring{p}) = 1$ (since $U(L(\mathring{p})) = U(1) = 1$) and thus we are left with

$$\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle = N(p)N^*(p') D^*_{\sigma\sigma'}(W(L^{-1}(p), p')) \delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}}).$$

Since we have both $\mathring{p}' = L^{-1}(p)p'$ and $\mathring{p} = L^{-1}(p)p$, we see that $\delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}})$ must be proportional to $\delta(\mathbf{p}' - \mathbf{p})$ (this is just the statement that $\delta(ax) = (1/|a|)\delta(x)$). Then the right side of the above equation vanishes for $\mathbf{p}' \neq \mathbf{p}$, and if $\mathbf{p}' = \mathbf{p}$ the Wigner rotation becomes simply

$$W(L^{-1}(p), p) = L^{-1}(L^{-1}(p)p) L^{-1}(p) L(p) = L^{-1}(\mathring{p}) = 1$$

because equation (158) applied to $\mathring{p}$ is just $\mathring{p} = L(\mathring{p})\mathring{p}$. But then

$$D^*_{\sigma\sigma'}(W(L^{-1}(p), p)) = \delta_{\sigma\sigma'}$$

and we can write

$$\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle = |N(p)|^2 \, \delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}}) \delta_{\sigma'\sigma}. \tag{167}$$

Lastly, we need to determine $N(p)$. Since p is related to $\mathring{p}$ by the same Lorentz transformation that relates p' to $\mathring{p}'$, we use equation (165) to write

$$\mathring{p}^0 \delta(\mathring{\mathbf{p}}' - \mathring{\mathbf{p}}) = p^0 \delta(\mathbf{p}' - \mathbf{p})$$

so that equation (167) becomes

$$\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle = |N(p)|^2 \, \frac{p^0}{\mathring{p}^0} \delta(\mathbf{p}' - \mathbf{p}) \delta_{\sigma'\sigma}.$$

Therefore, if we choose

$$N(p) = \sqrt{\mathring{p}^0 / p^0}$$

we are left with our final result

$$\langle \Psi_{p',\sigma'} | \Psi_{p,\sigma} \rangle = \delta(\mathbf{p}' - \mathbf{p}) \delta_{\sigma'\sigma}. \tag{168}$$

Now let us turn our attention to massless particles. Since there is no rest frame for the particle, we can't repeat the above procedure. In this case we take our standard 4-momentum to be the light-like 4-vector $\mathring{p}^\mu = (1, 0, 0, 1)$ so that $\mathring{p}^2 = 0$ and $\mathring{p}^0 > 0$. By definition we have $\mathring{\Lambda}\mathring{p} = \mathring{p}$, so for the timelike 4-vector $t^\mu = (1, 0, 0, 0)$ we have

$$(\mathring{\Lambda} t)^\mu (\mathring{\Lambda} t)_\mu = t^\mu t_\mu = 1 \tag{169a}$$

and

$$(\mathring{\Lambda} t)^\mu (\mathring{\Lambda} \mathring{p})_\mu = (\mathring{\Lambda} t)^\mu \mathring{p}_\mu = t^\mu \mathring{p}_\mu = 1. \tag{169b}$$

Any 4-vector $(\mathring{\Lambda} t)^\mu$ satisfying equation (169b) is of the form

$$(\mathring{\Lambda} t)^\mu = (1 + \zeta, \alpha, \beta, \zeta)$$

so equation (169a) yields the relation

$$\zeta = \frac{1}{2}(\alpha^2 + \beta^2). \tag{170}$$

Using equation (170), it is not hard to verify that

$$S(\alpha, \beta)^\mu{}_\nu = \begin{bmatrix} 1 + \zeta & \alpha & \beta & -\zeta \\ \alpha & 1 & 0 & -\alpha \\ \beta & 0 & 1 & -\beta \\ \zeta & \alpha & \beta & 1 - \zeta \end{bmatrix} \tag{171}$$

is a Lorentz transformation (i.e., $S^T gS = g$). Furthermore, it has the property that $(St)^\mu = (\mathring{\Lambda}t)^\mu$ and $(S\mathring{p})^\mu = (\mathring{\Lambda}\mathring{p})^\mu$. This *does not* mean that $S = \mathring{\Lambda}$, but it does mean that $S^{-1}\mathring{\Lambda}$ leaves the vector $t^\mu = (1, 0, 0, 0)$ invariant so it represents a pure rotation. In fact, it also leaves the vector $\mathring{p} = (1, 0, 0, 1)$ invariant, so it must be a rotation about the x^3-axis by some angle θ. In other words, we have

$$S^{-1}(\alpha, \beta)\mathring{\Lambda} = R(\theta)$$

where

$$R(\theta)^\mu{}_\nu = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{172}$$

Thus the most general element of the little group is of the form

$$\mathring{\Lambda}(\theta, \alpha, \beta) = S(\alpha, \beta)R(\theta). \tag{173}$$

Let us look at equation (173) for infinitesimal θ, α, β. To first order we have $\cos\theta = 1, \sin\theta = \theta$ and $\zeta = 0$, so multiplying the matrices from equations (171) and (172) and keeping first order terms yields

$$\mathring{\Lambda}(\theta, \alpha, \beta)^\mu{}_\nu = \begin{bmatrix} 1 & \alpha & \beta & 0 \\ \alpha & 1 & \theta & -\alpha \\ \beta & -\theta & 1 & -\beta \\ 0 & \alpha & \beta & 1 \end{bmatrix} := g^\mu{}_\nu + \mathring{\omega}^\mu{}_\nu$$

where

$$\mathring{\omega}_{\mu\nu} = g_{\mu\alpha}\mathring{\omega}^\alpha{}_\nu = \begin{bmatrix} 0 & \alpha & \beta & 0 \\ -\alpha & 0 & -\theta & \alpha \\ -\beta & \theta & 0 & \beta \\ 0 & -\alpha & -\beta & 0 \end{bmatrix}.$$

(Note this satisfies $\mathring{\omega}^\mu{}_\nu \mathring{p}^\nu = 0$ as required by equation (161).) The corresponding unitary transformation is

$$\begin{aligned} U(\mathring{\Lambda}(\theta, \alpha, \beta)) &= 1 + \frac{i}{2}\mathring{\omega}_{\mu\nu}M^{\mu\nu} = 1 + \frac{i}{2}(\mathring{\omega}_{0i}M^{0i} + \mathring{\omega}_{i0}M^{i0} + \mathring{\omega}_{ij}M^{ij}) \\ &= 1 + \frac{i}{2}\big(\alpha M^{01} + \beta M^{02} - \alpha M^{10} - \beta M^{20} \\ &\qquad - \theta M^{12} + \alpha M^{13} + \theta M^{21} + \beta M^{23} - \alpha M^{31} - \beta M^{32}\big). \end{aligned}$$

Using $M^{\mu\nu} = -M^{\nu\mu}$ along with the definitions

$$J_i = \frac{1}{2}\varepsilon_{ijk}M^{jk} \qquad \text{and} \qquad K_i = M_{i0} = -M^{i0}$$

(see equation (132)) we have

$$\begin{aligned} U(\mathring{\Lambda}(\theta,\alpha,\beta)) &= 1 + i[\alpha(-J_2 + K_1) + \beta(J_1 + K_2) - \theta J_3] \\ &:= 1 + i\alpha A + i\beta B - \theta J_3 \end{aligned} \tag{174}$$

where we have defined the Hermitian operators

$$A = -J_2 + K_1 \qquad \text{and} \qquad B = J_1 + K_2.$$

Using the commutation relations in equations (137) we find

$$\begin{aligned} [A, J_3] &= -iB \\ [B, J_3] &= +iA \\ [A, B] &= 0. \end{aligned}$$

Since A, B and $P_\mu P^\mu$ are commuting Hermitian operators, they may be simultaneously diagonalized by states $|\Psi_{\mathring{p},a,b}\rangle$ and we can write

$$\begin{aligned} A|\Psi_{\mathring{p},a,b}\rangle &= a|\Psi_{\mathring{p},a,b}\rangle \\ B|\Psi_{\mathring{p},a,b}\rangle &= b|\Psi_{\mathring{p},a,b}\rangle. \end{aligned}$$

But now this is a problem which we can see as follows. For proper rotations, **J** and **K** transform as ordinary vectors, and hence by equation (124) we have

$$U(R)J^i U^{-1}(R) = (R^{-1})^i{}_k J^k$$

with a similar result for **K**. (Here the matrix $R^i{}_j$ consists of the 3×3 spatial part of equation (172).) Then it is straightforward using equation (172) to show that

$$\begin{aligned} U[R(\theta)]A\, U^{-1}[R(\theta)] &= A\cos\theta - B\sin\theta \\ U[R(\theta)]B\, U^{-1}[R(\theta)] &= A\sin\theta + B\cos\theta. \end{aligned} \tag{175}$$

Defining the states

$$|\Psi^\theta_{\mathring{p},a,b}\rangle := U^{-1}[R(\theta)]|\Psi_{\mathring{p},a,b}\rangle$$

we then have

$$\begin{aligned} A|\Psi^\theta_{\mathring{p},a,b}\rangle &= A\big\{U^{-1}(R(\theta))|\Psi_{\mathring{p},a,b}\rangle\big\} \\ &= U^{-1}[R(\theta)]\big\{U[R(\theta)]A\, U^{-1}[R(\theta)]\big\}|\Psi_{\mathring{p},a,b}\rangle \\ &= (a\cos\theta - b\sin\theta)U^{-1}[R(\theta)]|\Psi_{\mathring{p},a,b}\rangle \\ &= (a\cos\theta - b\sin\theta)|\Psi^\theta_{\mathring{p},a,b}\rangle \end{aligned}$$

and similarly

$$B|\Psi^\theta_{\mathring{p},a,b}\rangle = (a\sin\theta + b\cos\theta)|\Psi^\theta_{\mathring{p},a,b}\rangle.$$

In other words, for arbitrary θ, the states $|\Psi^\theta_{\mathring{p},a,b}\rangle$ are simultaneous eigenstates of A and B. But massless particles are not observed to have any continuous degree of freedom like θ, and therefore we are forced to require that physical states be eigenvectors of A and B with eigenvalues $a = b = 0$.

Let us label our (physical) massless states $|\Psi_{\mathring{p},\sigma}\rangle$ so that $A|\Psi_{\mathring{p},\sigma}\rangle = B|\Psi_{\mathring{p},\sigma}\rangle = 0$, and where the label σ is the eigenvalue of the remaining generator J_3 in equation (174):

$$J_3|\Psi_{\mathring{p},\sigma}\rangle = \sigma|\Psi_{\mathring{p},\sigma}\rangle.$$

Since the spatial momentum $\mathring{\mathbf{p}}$ is in the 3-direction, this shows that σ is the component of angular momentum (spin) in the direction of motion. This is called the **helicity** of the particle.

11 The Dirac Equation Again

Now that we have studied direct product groups and representations of the Poincaré group, let us return to the Lorentz group decomposition (138):

$$\Lambda = e^{-i\mathbf{a}\cdot\mathbf{J}+i\mathbf{b}\cdot\mathbf{K}} = e^{-(i\mathbf{a}-\mathbf{b})\cdot\mathbf{A}-(i\mathbf{a}+\mathbf{b})\cdot\mathbf{B}} = e^{-(i\mathbf{a}-\mathbf{b})\cdot\mathbf{A}}e^{-(i\mathbf{a}+\mathbf{b})\cdot\mathbf{B}}$$

where

$$\mathbf{A} = \frac{1}{2}(\mathbf{J}+i\mathbf{K}) \qquad \text{and} \qquad \mathbf{B} = \frac{1}{2}(\mathbf{J}-i\mathbf{K})$$

and these commuting operators each satisfy the angular momentum commutation relations (141). This is exactly equivalent to the operator $SU(2)\otimes SU(2)$ that we described earlier, and hence we can label the representations by a pair of angular momentum states (j, j').

Let us label the $(2j+1)(2j'+1)$-dimensional representations

$$\Lambda = e^{-i(\mathbf{a}+i\mathbf{b})\cdot\mathbf{A}} \otimes e^{-i(\mathbf{a}-i\mathbf{b})\cdot\mathbf{B}} \tag{176}$$

by $D^{(j,j')}(\Lambda)$, where j labels the value of $\mathbf{A}^2$ and j' labels the value of $\mathbf{B}^2$. In particular, we consider the 2-dimensional representations

$$D(\Lambda) := D^{(1/2,0)}(\Lambda) \qquad \text{and} \qquad \overline{D}(\Lambda) := D^{(0,1/2)}(\Lambda).$$

Recall from the theory of spin 1/2 particles in quantum mechanics that the 2-dimensional representation of $SU(2)$ is just $\boldsymbol{\sigma}/2$ where the Pauli matrices are given by

$$\sigma_1 = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \qquad \sigma_2 = \begin{bmatrix} & -i \\ i & \end{bmatrix} \qquad \sigma_3 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}.$$

Then we see that

$$\text{For } D(\Lambda): \quad \mathbf{B} = 0 \implies \mathbf{J} = i\mathbf{K} \implies \mathbf{A} = i\mathbf{K} = \frac{\boldsymbol{\sigma}}{2}. \tag{177a}$$

$$\text{For } \overline{D}(\Lambda): \quad \mathbf{A} = 0 \implies \mathbf{J} = -i\mathbf{K} \implies \mathbf{B} = -i\mathbf{K} = \frac{\boldsymbol{\sigma}}{2}. \tag{177b}$$

Each of these 2-dimensional representations acts on a 2-component spinor that then transforms under either $D(\Lambda)$ or $\overline{D}(\Lambda)$ where

$$D(\Lambda) = e^{-i(\mathbf{a}+i\mathbf{b})\cdot\boldsymbol{\sigma}/2} \tag{178a}$$

and

$$\overline{D}(\Lambda) = e^{-i(\mathbf{a}-i\mathbf{b})\cdot\boldsymbol{\sigma}/2}. \tag{178b}$$

Note that $\mathbf{a}$ gives the rotation angle, and $\mathbf{b}$ gives the boost parameters.

It is important to understand that the representations D and $\overline{D}$ are *inequivalent* representations of the Lorentz group. In other words, there is no matrix S such that $S^{-1}D(\Lambda)S = \overline{D}(\Lambda)$ for all Λ. To see this, first note that using

$$e^{i\boldsymbol{\theta}\cdot\boldsymbol{\sigma}} = I\cos\theta + i(\widehat{\boldsymbol{\theta}}\cdot\boldsymbol{\sigma})\sin\theta = \begin{bmatrix} \cos\theta + i\widehat{\boldsymbol{\theta}}_3\sin\theta & i\widehat{\boldsymbol{\theta}}_-\sin\theta \\ i\widehat{\boldsymbol{\theta}}_+\sin\theta & \cos\theta - i\widehat{\boldsymbol{\theta}}_3\sin\theta \end{bmatrix}$$

(which follows from $(\boldsymbol{\theta}\cdot\boldsymbol{\sigma})^2 = \theta^2$ and where $\widehat{\boldsymbol{\theta}}_\pm = \widehat{\boldsymbol{\theta}}_1 \pm i\widehat{\boldsymbol{\theta}}_2$) it is easy to see that $\det D = \det\overline{D} = 1$. Thus D and $\overline{D}$ are 2×2 complex matrices with determinant equal to one. The set of all such matrices forms the group $SL(2,C)$. In general, a 2×2 complex matrix has eight independent components. But the requirement that it have unit determinant amounts to two equations relating the components, and hence an element of $SL(2,C)$ has six independent parameters. In the present case, these are the three boost parameters plus the three rotation angles.

Next, by explicit calculation you can easily show that

$$\sigma_2\boldsymbol{\sigma}^*\sigma_2 = -\boldsymbol{\sigma}.$$

Now define $S = i\sigma_2$ so that $S^{-1} = -i\sigma_2$. Then writing $D = e^{\mathbf{c}\cdot\boldsymbol{\sigma}}$ for simplicity we in fact have (using $(\sigma_2)^2 = 1$ and inserting this between products of $\boldsymbol{\sigma}^*$'s)

$$\begin{aligned} S^{-1}D^*S = \sigma_2 e^{\mathbf{c}^*\cdot\boldsymbol{\sigma}^*}\sigma_2 &= \sigma_2[1 + \mathbf{c}^*\cdot\boldsymbol{\sigma}^* + (1/2)(\mathbf{c}^*\cdot\boldsymbol{\sigma}^*)^2 + \cdots]\sigma_2 \\ &= 1 + \mathbf{c}^*\cdot(\sigma_2\boldsymbol{\sigma}^*\sigma_2) + (1/2)(\mathbf{c}^*\cdot(\sigma_2\boldsymbol{\sigma}^*\sigma_2))^2 + \cdots \\ &= 1 - \mathbf{c}^*\boldsymbol{\sigma} + (1/2)(-\mathbf{c}^*\cdot\boldsymbol{\sigma})^2 + \cdots \\ &= e^{-\mathbf{c}^*\cdot\boldsymbol{\sigma}} = \overline{D} \end{aligned}$$

since $-\mathbf{c}^* = -(-i(\mathbf{a}+i\mathbf{b})/2)^* = -i(\mathbf{a}-i\mathbf{b})/2$. In other words, $\overline{D}(\Lambda)$ is similar to $D(\Lambda)^*$ but not to $D(\Lambda)$. If $\overline{D}$ were equivalent to D, then this would mean that D was equivalent to D^*. That this cannot happen in general (in the present case of the 2-dimensional representation of $SL(2,C)$) can be shown by a counter example. Let M be a 2×2 matrix in $SL(2,C)$, and let its eigenvalues be λ and λ^{-1} where $|\lambda| \neq 1$ and $\operatorname{Im}\lambda \neq 0$. In other words, we can write

$$M = \begin{bmatrix} \lambda & \\ & 1/\lambda \end{bmatrix} \qquad \text{and} \qquad M^* = \begin{bmatrix} \lambda^* & \\ & 1/\lambda^* \end{bmatrix}.$$

But any matrix similar to M (i.e., equivalent) will have the same eigenvalues (since the characteristic polynomial doesn't change under a similarity transformation), and hence if M^* were equivalent to M we would have either $\lambda^* = \lambda$ or $\lambda^* = \lambda^{-1}$. Writing $\lambda = a + ib$ we have $\lambda^* = a - ib$ and

$$\lambda^{-1} = \frac{1}{a+ib} = \frac{a-ib}{|a|^2+|b|^2}.$$

Clearly λ^* is not equal to either λ or λ^{-1}, and thus M can't be similar to M^*. Therefore D can't be equivalent to $\overline{D}$ as claimed.

In the particular case of a pure boost $\Lambda = L(p)$, we have $\mathbf{a} = 0$ so that

$$D(L(p)) = e^{\mathbf{b}\cdot\boldsymbol{\sigma}/2} \qquad \text{and} \qquad \overline{D}(L(p)) = e^{-\mathbf{b}\cdot\boldsymbol{\sigma}/2}. \tag{179}$$

If the boost is along the z-direction, then $\mathbf{b} = u\hat{\mathbf{z}}$ and hence

$$D(L(p_3)) = e^{u\sigma_3/2} = \exp\begin{bmatrix} u/2 & 0 \\ 0 & -u/2 \end{bmatrix} = \begin{bmatrix} e^{u/2} & 0 \\ 0 & e^{-u/2} \end{bmatrix} \tag{180a}$$

and

$$\overline{D}(L(p_3)) = e^{-u\sigma_3/2} = \exp\begin{bmatrix} -u/2 & 0 \\ 0 & u/2 \end{bmatrix} = \begin{bmatrix} e^{-u/2} & 0 \\ 0 & e^{u/2} \end{bmatrix} \tag{180b}$$

where, by equations (53) and (54), the boost parameter u is defined by

$$\cosh u = \gamma \qquad \text{and} \qquad \sinh u = \gamma\beta.$$

Now what about the states of our system? In the last section we saw that the two Casimir operators $\mathscr{M}^2 = P_\mu P^\mu$ and $\mathscr{S}^2 = W_\mu W^\mu$ are the only operators that commute with all of the generators of Lorentz transformations, so our particles (the irreducible representations) can be specified by the Lorentz invariant eigenvalues m^2 and $s(s+1)$ of these operators. Since $[P_\mu, P_\nu] = [P_\mu, W_\nu] = 0$ while $[W_\mu, W_\nu] \neq 0$, we can label our states by the eigenvalues of the commuting operators $P_\mu, W_\nu W^\nu$ and one of the W_ν, say W_3. Thus we label our states by $|p\,\sigma\rangle$ where $p^2 = m^2$ (so the spatial components of p_μ are independent) and σ is the eigenvalue of W_3. (For simplicity we have suppressed the eigenvalue s of $\mathscr{S}^2$, i.e., we should really write $|p\,s\,\sigma\rangle$.)

Next, in the case of a pure rotation $\Lambda = R$ we have $\mathbf{b} = 0$ so that $R = e^{-i\mathbf{a}\cdot\mathbf{J}}$. From equations (178) we then see that $D(R) = \overline{D}(R)$. Applying this to the particular case of the Wigner rotation we then have $D(L^{-1}(\Lambda p)\Lambda L(p)) = \overline{D}(L^{-1}(\Lambda p)\Lambda L(p))$. Using the group property of the representations we can write this out as a matrix product:

$$D(L^{-1}(\Lambda p))D(\Lambda)D(L(p)) = \overline{D}(L^{-1}(\Lambda p))\overline{D}(\Lambda)\overline{D}(L(p)). \tag{181}$$

Let us define the two states

$$\begin{aligned} |p\,\sigma\,1\rangle &= \sum_{\sigma'} D^{-1}_{\sigma'\sigma}(L(p))|p\,\sigma'\rangle = \sum_{\sigma'} D_{\sigma'\sigma}(L^{-1}(p))|p\,\sigma'\rangle \\ |p\,\sigma\,2\rangle &= \sum_{\sigma'} \overline{D}^{-1}_{\sigma'\sigma}(L(p))|p\,\sigma'\rangle = \sum_{\sigma'} \overline{D}_{\sigma'\sigma}(L^{-1}(p))|p\,\sigma'\rangle \end{aligned} \tag{182}$$

where the sums are over the index values $-1/2$ and $1/2$. We can invert both of these to solve for $|p\,\sigma'\rangle$ and write

$$\sum_{\sigma''} D_{\sigma''\sigma'}(L(p))|p\,\sigma''\,1\rangle = |p\,\sigma'\rangle = \sum_{\sigma''} \overline{D}_{\sigma''\sigma'}(L(p))|p\,\sigma''\,2\rangle. \tag{183}$$

Now put these back into the right side of equations (182) to obtain the coupled equations

$$\begin{aligned} |p\,\sigma\,1\rangle &= \sum_{\sigma'\sigma''} |p\,\sigma''\,2\rangle \overline{D}_{\sigma''\sigma'}(L(p)) D_{\sigma'\sigma}(L^{-1}(p)) \\ |p\,\sigma\,2\rangle &= \sum_{\sigma'\sigma''} |p\,\sigma''\,1\rangle D_{\sigma''\sigma'}(L(p)) \overline{D}_{\sigma'\sigma}(L^{-1}(p)). \end{aligned} \tag{184}$$

To evaluate these matrix products, consider a boost in the z-direction. From equations (180) along with $D(L^{-1}) = D^{-1}(L)$ we have

$$\begin{aligned} \overline{D}(L(p_3))D^{-1}(L(p_3)) &= \begin{bmatrix} e^{-u/2} & 0 \\ 0 & e^{u/2} \end{bmatrix} \begin{bmatrix} e^{-u/2} & 0 \\ 0 & e^{u/2} \end{bmatrix} = \begin{bmatrix} e^{-u} & 0 \\ 0 & e^{u} \end{bmatrix} \\ D(L(p_3))\overline{D}^{-1}(L(p_3)) &= \begin{bmatrix} e^{u/2} & 0 \\ 0 & e^{-u/2} \end{bmatrix} \begin{bmatrix} e^{u/2} & 0 \\ 0 & e^{-u/2} \end{bmatrix} = \begin{bmatrix} e^{u} & 0 \\ 0 & e^{-u} \end{bmatrix}. \end{aligned} \tag{185}$$

Now note the identities

$$\begin{aligned} e^{u} &= \cosh u + \sinh u = \gamma + \gamma\beta = \frac{p_0}{m} + \frac{p_3}{m} \\ e^{-u} &= \cosh u - \sinh u = \gamma - \gamma\beta = \frac{p_0}{m} - \frac{p_3}{m} \end{aligned}$$

where we also used equations (153). Then we have

$$\begin{aligned} \begin{bmatrix} e^{-u} & 0 \\ 0 & e^{u} \end{bmatrix} &= \begin{bmatrix} p_0/m - p_3/m & \\ & p_0/m + p_3/m \end{bmatrix} = \frac{p_0}{m}I - \frac{p_3}{m}\sigma_3 = \frac{p_0}{m}I - \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m} \\ \begin{bmatrix} e^{u} & 0 \\ 0 & e^{-u} \end{bmatrix} &= \begin{bmatrix} p_0/m + p_3/m & \\ & p_0/m - p_3/m \end{bmatrix} = \frac{p_0}{m}I + \frac{p_3}{m}\sigma_3 = \frac{p_0}{m}I + \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m}. \end{aligned} \tag{186}$$

Combining equations (184), (185) and (186) yields

$$\begin{aligned} |p\,\sigma\,1\rangle &= \sum_{\sigma'} |p\,\sigma'\,2\rangle \left(\frac{p_0}{m}I - \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m}\right)_{\sigma'\sigma} \\ |p\,\sigma\,2\rangle &= \sum_{\sigma'} |p\,\sigma'\,1\rangle \left(\frac{p_0}{m}I + \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m}\right)_{\sigma'\sigma} \end{aligned} \tag{187}$$

If we write

$$|p\,1\rangle = \begin{bmatrix} |p\,\frac{1}{2}\,1\rangle \\ |p\,-\frac{1}{2}\,1\rangle \end{bmatrix}$$

so that $|p\,1\rangle^T = [\,|p\,\frac{1}{2}\,1\rangle \quad |p\,-\frac{1}{2}\,1\rangle\,]$ with a similar result for $|p\,2\rangle$, then equations (187) can be written in matrix form as

$$|p\,1\rangle^T = |p\,2\rangle^T\left(\frac{p_0}{m}I - \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m}\right) \qquad \text{and} \qquad |p\,2\rangle^T = |p\,1\rangle^T\left(\frac{p_0}{m}I + \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m}\right)$$

where the first row and column correspond to $\sigma = 1/2$ and the second row and column correspond to $\sigma = -1/2$. Rearranging, these become

$$-m|p\,1\rangle^T + |p\,2\rangle^T(p_0 I - \mathbf{p}\cdot\boldsymbol{\sigma}) = 0$$

$$|p\,1\rangle^T(p_0 I + \mathbf{p}\cdot\boldsymbol{\sigma}) - m|p\,2\rangle^T = 0$$

which can be combined into the form

$$[\,|p\,1\rangle^T \quad |p\,2\rangle^T\,]\begin{bmatrix} -mI & p_0 + \mathbf{p}\cdot\boldsymbol{\sigma} \\ p_0 I - \mathbf{p}\cdot\boldsymbol{\sigma} & -mI \end{bmatrix} = 0.$$

Defining the row vector

$$\big(\,|p\,\zeta\rangle\big)^T = \big[\,|p\,\tfrac{1}{2}\,1\rangle \quad |p\,-\tfrac{1}{2}\,1\rangle \quad |p\,\tfrac{1}{2}\,2\rangle \quad |p\,-\tfrac{1}{2}\,2\rangle\,\big]$$

(where $\zeta = 1, \ldots, 4$) we have

$$\sum_\zeta |p\,\zeta\rangle \begin{bmatrix} -mI & p_0 + \mathbf{p}\cdot\boldsymbol{\sigma} \\ p_0 I - \mathbf{p}\cdot\boldsymbol{\sigma} & -mI \end{bmatrix}_{\zeta\zeta'} = 0. \tag{188}$$

Let $|u\rangle$ denote an arbitrary state and let $\langle p\,\zeta|u\rangle = u_\zeta(p)$. Noting $\langle u|p\,\zeta\rangle = \langle p\,\zeta|u\rangle^*$ and using the summation convention on the index ζ, equation (188) becomes

$$\begin{aligned} &\begin{bmatrix} -mI & p_0 + \mathbf{p}\cdot\boldsymbol{\sigma} \\ p_0 I - \mathbf{p}\cdot\boldsymbol{\sigma} & -mI \end{bmatrix}^T_{\zeta'\zeta} u^*_\zeta(p) \\ &\qquad = \left[p_0\begin{pmatrix} & 1 \\ 1 & \end{pmatrix} + \mathbf{p}\cdot\begin{pmatrix} & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & \end{pmatrix} - m\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}\right]^T_{\zeta'\zeta} u^*_\zeta(p) = 0. \end{aligned} \tag{189}$$

Now define

$$\gamma^0 = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\gamma} = \begin{pmatrix} & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & \end{pmatrix} \tag{190}$$

and note $\gamma^{\mu\dagger} = \gamma^0\gamma^\mu\gamma^0$ so that $\gamma^{0\dagger} = \gamma^0$ and $\gamma^{i\dagger} = -\gamma^i$. (The representation (190) is called the **chiral** (or **Weyl**) representation of the Dirac matrices, but be aware that different authors use different sign conventions.) Taking the complex conjugate of equation (189) yields

$$0 = (p_0\gamma^0 + \mathbf{p}\cdot\boldsymbol{\gamma} - mI)^\dagger u(p) = (p_0\gamma^{0\dagger} + \mathbf{p}\cdot\boldsymbol{\gamma}^\dagger - mI)u(p)$$

or finally

$$(p_0\gamma^0 - \mathbf{p}\cdot\boldsymbol{\gamma} - m)u(p) = 0$$

i.e.,

$$(\gamma^\mu p_\mu - m)u(p) = 0$$

which is the Dirac equation.

Let us write out the Dirac equation in the chiral basis as

$$\begin{bmatrix} -m & p_0 - \boldsymbol{\sigma}\cdot\mathbf{p} \\ p_0 + \boldsymbol{\sigma}\cdot\mathbf{p} & -m \end{bmatrix} \begin{bmatrix} \psi_L \\ \psi_R \end{bmatrix} = 0$$

where we have written the 4-component Dirac spinor as a combination of two 2-component spinors, called **Weyl spinors**. We can put this into an even more concise form by defining

$$\sigma^\mu = (I, \boldsymbol{\sigma}) \qquad \text{and} \qquad \overline{\sigma}^\mu = (I, -\boldsymbol{\sigma})$$

in which case we have

$$\gamma^\mu = \begin{bmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{bmatrix}$$

and the Dirac equation becomes

$$\begin{bmatrix} -m & p_\mu\sigma^\mu \\ p_\mu\overline{\sigma}^\mu & -m \end{bmatrix} \begin{bmatrix} \psi_L \\ \psi_R \end{bmatrix} = 0.$$

In the particular case of massless particles this becomes what are known as the **Weyl equations**:

$$(p_0 - \boldsymbol{\sigma}\cdot\mathbf{p})\psi_R = 0 \qquad \text{and} \qquad (p_0 + \boldsymbol{\sigma}\cdot\mathbf{p})\psi_L = 0.$$

Since in the massless case we have $p_0 = |\mathbf{p}|$, these can be written

$$\boldsymbol{\sigma}\cdot\widehat{\mathbf{p}}\psi_L = -\psi_L \qquad \text{and} \qquad \boldsymbol{\sigma}\cdot\widehat{\mathbf{p}}\psi_R = \psi_R$$

which shows that ψ_L has helicity -1 and ψ_R has helicity $+1$. Therefore we call ψ_L a **left-handed spinor** and ψ_R a **right-handed spinor**. Thus the Weyl spinors are helicity eigenstates, which is the reason for choosing the Weyl representation. Note also that in the Weyl representation we have

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}$$

so that

$$\frac{1}{2}(1+\gamma_5) = \begin{bmatrix} 0 & \\ & 1 \end{bmatrix} \qquad \text{and} \qquad \frac{1}{2}(1-\gamma_5) = \begin{bmatrix} 1 & \\ & 0 \end{bmatrix}.$$

Applying these operators to the 4-component spinor we see that

$$\frac{1}{2}(1+\gamma_5) \begin{bmatrix} \psi_L \\ \psi_R \end{bmatrix} = \begin{bmatrix} 0 \\ \psi_R \end{bmatrix} \qquad \text{and} \qquad \frac{1}{2}(1-\gamma_5) \begin{bmatrix} \psi_L \\ \psi_R \end{bmatrix} = \begin{bmatrix} \psi_L \\ 0 \end{bmatrix}$$

and hence the operators $(1\pm\gamma_5)/2$ project out the right- and left-handed spinors. I leave it as an easy exercise to show that an interaction term in the Lagrangian of the form

$$\overline{\psi}\gamma^\mu \left(\frac{1-\gamma_5}{2}\right)\psi$$

results in a current containing only left-handed spinors. Since we know that $\overline{\psi}\gamma^\mu\psi$ transforms as a vector and $\overline{\psi}\gamma^\mu\gamma_5\psi$ transforms as an axial vector, this is called a "V − A interaction" (read "V minus A"), and is fundamental to the description of the weak interactions.