Physics 221AB
Spring 1997
Notes 36

Lorentz Transformations in Quantum Mechanics and the Covariance of the Dirac Equation

These notes supplement Chapter 2 of Bjorken and Drell, which concerns the covariance of the Dirac equation under Lorentz transformations. The presentation of Bjorken and Drell is not easy to follow in certain respects, and these notes will fill in some of the missing steps as well as correct some errors. These notes go into more detail than is either in Bjorken and Drell or the lectures, particularly on the two 2-dimensional representations $S(\Lambda)$ and $\tilde{S}(\Lambda)$ of the Lorentz group. In the lecture we will "discover" these two representations by reworking the presentation of Bjorken and Drell, and working with the Weyl representation of the Dirac algebra.

We begin with some conventions and notation regarding Lorentz transformations. We let the 4-dimensional space-time contravariant coordinate vector be

$$x^\mu = (ct, \mathbf{x}) = (x^0, x^1, x^2, x^3), \tag{36.1}$$

and we let the metric tensor be

$$g_{\mu\nu} = g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{36.2}$$

The metric tensor is used to raise and lower indices in the usual way; for example, we have

$$x_\mu = g_{\mu\nu}\, x^\nu = (ct, -\mathbf{x}). \tag{36.3}$$

Likewise, the momentum 4-vector is

$$p^\mu = (\frac{E}{c}, \mathbf{p}), \qquad p_\mu = (\frac{E}{c}, -\mathbf{p}), \tag{36.4}$$

so that

$$p^\mu p_\mu = \frac{E^2}{c^2} - p^2 = m^2c^2. \tag{36.5}$$

In quantum mechanics, the momentum 4-vector is associated with a 4-vector of operators,

$$p^\mu = i\hbar\frac{\partial}{\partial x_\mu} = \Bigl(i\frac{\hbar}{c}\frac{\partial}{\partial t}, -i\hbar\nabla\Bigr) = \Bigl(\frac{H}{c}, \mathbf{p}\Bigr). \tag{36.6}$$

Now let X^μ be a generic contravariant vector. The scalar product of this vector with itself is

$$X^\mu X_\mu = X^\mu g_{\mu\nu} X^\nu. \tag{36.7}$$

We consider a linear transformation of this vector,

$$X'^\mu = \Lambda^\mu{}_\nu X^\nu, \tag{36.8}$$

specified by the 4×4 matrix $\Lambda^\mu{}_\nu$, and demand that it preserve the Minkowski scalar product,

$$X'^\mu g_{\mu\nu} X'^\nu = X^\mu g_{\mu\nu} X^\nu \tag{36.9}$$

for all vectors X^μ. This imposes a condition on the coefficients of the linear transformation,

$$\Lambda^\mu{}_\alpha \, g_{\mu\nu} \, \Lambda^\nu{}_\beta = g_{\alpha\beta}, \tag{36.10}$$

which by definition qualifies the transformation (36.8) as a *Lorentz transformation*. The physical origin of Eq. (36.10) in the constancy of the velocity of light, independent of the state of motion of the observer, and the usual mechanical and electromagnetic consequences of special relativity are discussed by Jackson, and will not be repeated here. We simply comment that if we identify the vector X^μ with the space-time coordinate vector $x^\mu = (ct, \mathbf{x})$, then the Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ is a mapping of space-time onto itself which preserves the Lorentz or Minkowski scalar product $x^\mu x_\mu = c^2t^2 - |\mathbf{x}|^2$. Thus, the character of a 4-vector, space-like, time-like or light-like ($x^\mu x_\mu < 0$, > 0, or $= 0$, respectively) is preserved by the transformation. In particular, the light cone is mapped onto itself.

It is convenient to write Eq. (36.10) in matrix form. We let $\mathsf{\Lambda}$ be the matrix whose components are $\Lambda^\mu{}_\nu$, with the first index being contravariant and the second covariant; it is necessary to specify this, because by raising and lowering indices, we can create three other component matrices, $\Lambda^{\mu\nu}$, $\Lambda_{\mu\nu}$, and $\Lambda_\mu{}^\nu$ (represented by matrices $\mathsf{\Lambda g}$, $\mathsf{g\Lambda}$ and $\mathsf{g\Lambda g}$, respectively). Then in matrix form Eq. (36.10) becomes

$$\mathsf{\Lambda}^t \mathsf{g} \mathsf{\Lambda} = \mathsf{g}. \tag{36.11}$$

This condition may be compared with the usual condition defining an orthogonal matrix,

$$\mathsf{R}^t \, \mathsf{I} \, \mathsf{R} = \mathsf{I}, \tag{36.12}$$

which shows that the matrices $\mathsf{\Lambda}$ defining a Lorentz transformation are orthogonal in a generalized sense, being taken with respect to the indefinite metric g instead of the Euclidean metric I.

We now explore the properties of the matrices Λ. Since $\det \mathsf{g} = -1$, it follows immediately from Eq. (36.11) that

$$\det \Lambda = \pm 1. \tag{36.13}$$

From this it follows that Lorentz transformations are always invertible. In fact, by multiplying Eq (36.11) through by g and using $\mathsf{g}^2 = \mathsf{I}$, we find

$$\Lambda^{-1} = \mathsf{g}\Lambda^t \mathsf{g}, \tag{36.14}$$

or, in components,

$$(\Lambda^{-1})^\mu{}_\nu = g^{\mu\alpha}\, \Lambda^\beta{}_\alpha\, g_{\beta\nu} = \Lambda_\nu{}^\mu. \tag{36.15}$$

It is easy to show that if Λ, Λ_1, and Λ_2 are Lorentz transformations, then so are Λ^{-1}, Λ^t, and $\Lambda_1\Lambda_2$. Thus, the Lorentz transformations form a group, sometimes denoted $O(3,1)$, to indicate matrices which are orthgogonal with respect to a metric with one time-like and three space-like dimensions. The fact that Λ^t is a Lorentz transformation goes beyond the group property. We will call the group $O(3,1)$ the *full Lorentz group*, to distinguish it from the proper Lorentz group to be introduced shortly. The Lorentz group is a 6-parameter group, which can be seen from the fact that an arbitrary 4×4 matrix has 16 independent components, while the condition (36.11), containing a symmetric matrix on both sides of the equation, constitutes 10 constraints. Physically, the six parameters identifying a Lorentz transformation can be identified with three Euler angles specifying a rotation, plus the three components of a velocity specifying a boost.

The full Lorentz group $O(3,1)$ is a 6-dimensional manifold, which can be thought of as a 6-dimensional surface living in the 16-dimensional space of real 4×4 matrices. This group manifold consists of four disconnected pieces [recall that the group manifold for $O(3)$ consists of two disconnected pieces, consisting of the proper and improper rotations]. The four pieces are identified by the sign of $\Lambda^0{}_0$, and by the sign of $\det \Lambda$. As for $\Lambda^0{}_0$, this coefficient of the Lorentz transformation connects the old and new time coordinates t and t', and apart from sign, is the usual relativistic factor of time dilation, $\gamma = 1/\sqrt{1-(v/c)^2}$, which is always ≥ 1. In fact, it is a simple consequence of the definition (36.11) that

$$|\Lambda^0{}_0| \geq 1. \tag{36.16}$$

Negative values of $\Lambda^0{}_0$ indicate Lorentz transformations which reverse the direction of time, thereby mapping the forward light cone into the backward light cone; and positive values indicate Lorentz transformations which preserve the direction of time. Thus, we have either $-\infty < \Lambda^0{}_0 \leq -1$, or $1 \leq \Lambda^0{}_0 < \infty$.

We will now characterize the four pieces or subsets of the full Lorentz group, and give examples of matrices from each subset. Matrices in the first subset satisfy ${\Lambda^0}_0 \geq 1$ and $\det \Lambda = +1$, and are called *proper Lorentz transformations.* The subset of proper transformations forms a group by itself, which we will call the *proper Lorentz group.* An example of a Lorentz transformation in this subset is the identity,

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{36.17}$$

and all proper Lorentz transformations can be continuously connected with the identity. Matrices in the second subset satisfy ${\Lambda^0}_0 \geq 1$ and $\det \Lambda = -1$. This subset contains the parity transformation,

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{36.18}$$

which reverses the spatial components but leaves the time component alone. Matrices in the third subset satisfy ${\Lambda^0}_0 \leq -1$ and $\det \Lambda = -1$. This subset contains the time-reversal transformation,

$$\Lambda = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{36.19}$$

which reverses the direction of time but leaves the spatial components alone. Finally, matrices in the fourth subset satisfy ${\Lambda^0}_0 \leq 1$ and $\det \Lambda = +1$, and this subset contains the total inversion operation,

$$\Lambda = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{36.20}$$

which is the product of parity and time-reversal. None of the last three subsets forms a group, since none contains the identity element. However, an arbitrary element of the second subset can be written as the matrix in Eq. (36.18) times a proper Lorentz transformation, and similar statements can be made about the third and fourth subset.

For most of the following discussion, we will primarily be interested in the proper Lorentz transformations, which consists of rotations and boosts and their products which can be continuously connected with the identity. We split off the discrete symmetries of parity and time-reversal, and deal with them separately later.

In the following discussion, we will generally interpret Lorentz transformations in a active sense. For example, when we transform a momentum 4-vector according to

$$p'^\mu = \Lambda^\mu{}_\nu p^\nu, \tag{36.21}$$

we will interpret p^μ and p'^μ as momentum vectors representing two states of motion of a given particle, whose components are measured or computed in a given Lorentz frame. We may think of p^μ as the old 4-momentum of a particle, and p'^μ as the new 4-momentum, after some rotation or boost operation has been applied to the particle. This contrasts with the (more usual) passive interpretation, in which p^μ and p'^μ represent the same state of motion of a particle, as viewed in two Lorentz frames. Recall that we took the active point of view earlier in our discussion of spatial rotations, and now we extend this to Lorentz transformations. Bjorken and Drell and most books on special relativity take the passive point of view, but we adopted the active point of view in our earlier discussion of rotations (see Notes 9), and we will continue it here.

We consider now *elementary Lorentz transformations*, which we define to be proper Lorentz transformations representing pure rotations about one of the coordinate axes or pure boosts along one of the coordinate axes. The three elementary rotations are

$$\begin{aligned}
\mathsf{\Lambda} = \mathsf{R}(\hat{\mathbf{x}}, \theta) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix}, \\
\mathsf{\Lambda} = \mathsf{R}(\hat{\mathbf{y}}, \theta) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & 0 & \sin\theta \\ 0 & 0 & 1 & 0 \\ 0 & -\sin\theta & 0 & \cos\theta \end{pmatrix}, \\
\mathsf{\Lambda} = \mathsf{R}(\hat{\mathbf{z}}, \theta) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},
\end{aligned} \tag{36.22}$$

which are just as in Eq. (9.13) except that the matrices are now 4-dimensional to accomodate the $\mu = 0$ row and column. As for the boosts, we parameterize them by means of the *rapidity* parameter λ, defined in terms of the velocity by

$$\tanh\lambda = \frac{v}{c} = \beta, \tag{36.23}$$

where v is the velocity of the boost and $\beta = v/c$ is the usual parameter of relativity theory. Thus, when the velocity ranges over $-c < v < +c$, the rapidity ranges over $-\infty < \lambda < +\infty$.

Also, we have the relations,

$$\begin{aligned}\sinh\lambda &= \frac{v/c}{\sqrt{1-(v/c)^2}} = \beta\gamma,\\ \cosh\lambda &= \frac{1}{\sqrt{1-(v/c)^2}} = \gamma,\end{aligned} \tag{36.24}$$

where $\gamma = 1/\sqrt{1-\beta^2}$ is the usual parameter of relativity theory. In terms of the rapidity, the three elementary boosts are

$$\begin{aligned}\Lambda = \mathsf{B}(\hat{\mathbf{x}},\lambda) &= \begin{pmatrix} \cosh\lambda & \sinh\lambda & 0 & 0\\ \sinh\lambda & \cosh\lambda & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix},\\ \Lambda = \mathsf{B}(\hat{\mathbf{y}},\lambda) &= \begin{pmatrix} \cosh\lambda & 0 & \sinh\lambda & 0\\ 0 & 1 & 0 & 0\\ \sinh\lambda & 0 & \cosh\lambda & 0\\ 0 & 0 & 0 & 1\end{pmatrix},\\ \Lambda = \mathsf{B}(\hat{\mathbf{z}},\lambda) &= \begin{pmatrix} \cosh\lambda & 0 & 0 & \sinh\lambda\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ \sinh\lambda & 0 & 0 & \cosh\lambda\end{pmatrix}.\end{aligned} \tag{36.25}$$

To explain where these matrices representing elementary rotations and boosts came from, we must elaborate on what the active interpretation of the Lorentz transformations means. In the case of the rotations, the active meaning is contained in a geometrical picture of vectors before and after the rotation, as discussed in class and in Notes 9. In the case of the boosts, an example will illustrate the active meaning. Consider the transformation of a momentum vector as in Eq. (36.21). Let $p^\mu = (mc, 0)$ be the 4-momentum of a particle at rest, and let Λ represent a boost along the x-axis. Then by direct matrix multiplication by the matrix $\mathsf{B}(\hat{\mathbf{x}},\lambda)$ in Eq. (36.25) we find $p'^\mu = (E'/c, \mathbf{p}')$, where

$$\begin{aligned}E' &= \gamma mc^2,\\ p'^1 &= \gamma mv,\\ p'^2 &= p'^3 = 0.\end{aligned} \tag{36.26}$$

The final momentum 4-vector p'^μ is that of a particle moving in the positive x-direction (assuming $v > 0$), and this fact justifies the form given in Eq. (36.25) for the boost $\mathsf{B}(\hat{\mathbf{x}},\lambda)$.

An arbitrary proper Lorentz transformation can be written as a product of elementary Lorentz transformations (about or along the coordinate axes). This is a theorem (which we shall not prove) which is obviously a generalization of the representation of arbitrary rotations in terms of Euler angles. It is also a theorem that an arbitrary proper Lorentz transformation can be uniquely represented as a product of a pure rotation (about some axis) times a pure boost (along some, generally different, axis). The pure rotations alone form a subgroup of the proper Lorentz group, which is obviously the group $SO(3)$ of proper rotations; but the pure boosts do not form a group, as the product of two boosts is in general not another boost. In fact, the commutator of two infinitesimal boosts is an infinitesimal rotation, a fact which lies at the heart of Thomas precession.

Pure rotations and pure boosts (along an arbitrary axis) can be written in exponential form,

$$\mathsf{R}(\hat{\mathbf{n}},\theta) = \exp(\theta\hat{\mathbf{n}}\cdot\mathbf{J}), \tag{36.27a}$$

$$\mathsf{B}(\hat{\mathbf{m}},\lambda) = \exp(\lambda\hat{\mathbf{m}}\cdot\mathbf{K}), \tag{36.27b}$$

where $\mathbf{J}$ and $\mathbf{K}$ are 3-vectors of 4×4 matrices. As we say, the matrices $\mathbf{J}$ generate the rotations, and the matrices $\mathbf{K}$ generate the boosts. The matrices $\mathbf{J}$ are defined by

$$\mathsf{J}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mathsf{J}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix},$$

$$\mathsf{J}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{36.28}$$

which are the same as the matrices $\mathbf{J}$ defined by Eq. (9.17), except for the expansion to four dimensions to include the time component. In the case of the boosts, the vector $\hat{\mathbf{m}}$ in Eq. (36.27b) is the direction of the boost, and λ is the rapidity. The matrices $\mathbf{K}$ are defined by

$$\mathsf{K}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathsf{K}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathsf{K}_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{36.29}$$

It is easily verified that with these definitions of the $\mathbf{K}$ matrices, and with $\hat{\mathbf{m}}$ in Eq. (36.27) chosen to be $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, or $\hat{\mathbf{z}}$, Eqs. (36.25) are reproduced.

For applications to the Dirac equation, we will need spinor representations of the Lorentz transformations, i.e., matrices $S(\Lambda)$ which are parameterized by the classical Lorentz transformations Λ and which satisfy the group representation law,

$$S(\Lambda_1)S(\Lambda_2) = S(\Lambda_1\Lambda_2). \tag{36.30}$$

This is exactly as in Eq. (10.3), which refers to the spinor representations of ordinary rotations. Equation (36.30) implies

$$S(\mathsf{I}) = 1, \tag{36.31}$$

where I is the 4×4 identity matrix, and 1 is the identity matrix of the representation, which has the same dimensionality as S. Equation (36.30) also implies

$$S(\Lambda^{-1}) = S(\Lambda)^{-1}. \tag{36.32}$$

Actually, as we will see, when we finally obtain explicit formulas for the matrices S, we will find that they form a double-valued representation of the Lorentz transformations, just as the spinor rotation group $SU(2)$ is a double-valued representation of the proper rotation group $SO(3)$. That is, in the end we will find that Eqs. (36.30)–(36.32) are only satisfied to within a sign. Nevertheless, we can proceed as if we are seeking a genuine (single-valued) representation $S(\Lambda)$, and see what happens.

Just as we did in Notes 10 for the case of spinor rotations, our strategy for finding representations of the Lorentz transformation will be first to find representations of infinitesimal transformations, which can be expressed in terms of the generators $\mathbf{J}$ and $\mathbf{K}$ of the Lorentz group. These generators span the Lie algebra of the Lorentz group. Then from infinitesimal transformations, we can build up finite ones, both for the classical Lorentz transformations Λ and for the representations $S(\Lambda)$. We begin by exploring the Lie algebra of the Lorentz group.

First we consider infinitesimal pure rotations or pure boosts, which can be obtained from Eqs. (36.27) by taking θ or λ as small, so that

$$\mathsf{R}(\hat{\mathbf{n}},\theta) = \exp(\theta\hat{\mathbf{n}}\cdot\mathbf{J}) = \mathsf{I} + \theta\hat{\mathbf{n}}\cdot\mathbf{J}, \tag{36.33a}$$

$$\mathsf{B}(\hat{\mathbf{m}},\lambda) = \exp(\lambda\hat{\mathbf{m}}\cdot\mathbf{K}) = \mathsf{I} + \lambda\hat{\mathbf{m}}\cdot\mathbf{K}. \tag{36.33b}$$

An arbitrary infinitesimal Lorentz transformation is a product of an infinitesimal rotation times an infinitesimal boost, which we write in the form

$$\Lambda = \mathsf{I} + \epsilon(\mathbf{a}\cdot\mathbf{J} + \mathbf{b}\cdot\mathbf{K}), \tag{36.34}$$

where ϵ is a reminder that the correction term is small, and where we have set

$$\epsilon\mathbf{a} = \theta\hat{\mathbf{n}}, \qquad \epsilon\mathbf{b} = \lambda\hat{\mathbf{m}}. \tag{36.35}$$

We see that an arbitrary infinitesimal Lorentz transformation is specified by two real vectors $\mathbf{a}$ and $\mathbf{b}$, which make six parameters altogether. Obviously, $\epsilon\mathbf{a}$ represents the axis and angle of the infinitesimal rotation, and $\epsilon\mathbf{b}$ represents the axis and velocity of the infinitesimal boost.

Sometimes it is convenient to write infinitesimal Lorentz transformations in another way, namely,

$$\Lambda^\mu{}_\nu = \delta^\mu_\nu + \epsilon\Omega^\mu{}_\nu, \tag{36.36}$$

where $\Omega^\mu{}_\nu$ is the correction term and where ϵ is a reminder that this term is small. We will also write this in matrix form,

$$\Lambda = \mathsf{I} + \epsilon\Omega, \tag{36.37}$$

where Ω is the matrix with components $\Omega^\mu{}_\nu$. Comparing Eq. (36.37) with Eq. (36.34), we obviously have

$$\Omega = \mathbf{a}\cdot\mathbf{J} + \mathbf{b}\cdot\mathbf{K}. \tag{36.38}$$

Substituting Eq. (36.37) into Eq. (36.11) and expanding to first order in ϵ, we find

$$\Omega^t \mathsf{g} + \mathsf{g}\Omega = 0, \tag{36.39}$$

which is equivalent to the statement that $\mathsf{g}\Omega$ is antisymmetric. Since $\mathsf{g}\Omega$ has components $g_{\mu\alpha}\Omega^\alpha{}_\nu = \Omega_{\mu\nu}$, Eq. (36.39) is equivalent to

$$\Omega_{\mu\nu} = -\Omega_{\nu\mu}. \tag{36.40}$$

If we explicitly write out Ω according to Eq. (36.38) and lower the first index, we find

$$\Omega_{\mu\nu} = \begin{pmatrix} 0 & b_1 & b_1 & b_3 \\ -b_1 & 0 & a_3 & -a_2 \\ -b_2 & -a_3 & 0 & a_1 \\ -b_3 & a_2 & -a_1 & 0 \end{pmatrix}. \tag{36.41}$$

Thus, an arbitrary infinitesimal Lorentz transformation is specified either by two 3-vectors $\mathbf{a}$ and $\mathbf{b}$ or by the antisymmetric 4×4 tensor $\Omega_{\mu\nu}$. At this point you may want to look at the discussion surrounding Eq. (9.16), which concerned infinitesimal (ordinary) rotations.

What is the form of the representation $S(\mathsf{\Lambda})$ of the infinitesimal Lorentz transformation (36.34)? It must of course also be infinitesimal, and the correction terms must involve linear combinations of the same coefficients $\mathbf{a}$ and $\mathbf{b}$ which appear in Eq. (36.34), if the group representation law (36.30) is to hold. That is, we have the parallel equations,

$$\mathsf{\Lambda} = \mathsf{I} + \epsilon(\mathbf{a}\cdot\mathsf{\mathbf{J}} + \mathbf{b}\cdot\mathsf{\mathbf{K}}), \tag{36.42a}$$

$$S(\mathsf{\Lambda}) = 1 - i\epsilon(\mathbf{a}\cdot\mathbf{J} + \mathbf{b}\cdot\mathbf{K}), \tag{36.42b}$$

which imply one another, where $\mathbf{J}$ and $\mathbf{K}$ are two new vectors of matrices of the same dimensionality as S. These matrices are as yet undetermined. The factor of $-i$ in Eq. (36.42b) is just a convention of quantum mechanics, where we prefer to see Hermitian generators generate unitary operators. We might also have included a conventional factor of $1/\hbar$ in Eq. (36.42b), but in these notes we will absorb the $\hbar$ into the definition of $\mathbf{J}$ and $\mathbf{K}$ (or equivalently, choose units such that $\hbar = 1$).

How do we determine the matrices $\mathbf{J}$ and $\mathbf{K}$? First, these matrices (multiplied by $-i$) must satisfy the same commutation relations as the matrices $\mathsf{\mathbf{J}}$ and $\mathsf{\mathbf{K}}$, in order that the group representation law (36.30) should hold for finite transformations. As discussed in Notes 10, the commutation relations of a Lie algebra contain all the information needed to construct finite group operations, and this is why so much attention is devoted to the commutation relations of the generators. Let us therefore work out the commutation relations of the generators $\mathsf{\mathbf{J}}$ and $\mathsf{\mathbf{K}}$. A direct calculation of commutators, based on Eqs. (36.28) and (36.29), yields

$$\begin{aligned}
[\mathsf{J}_i, \mathsf{J}_j] &= \epsilon_{ijk}\, \mathsf{J}_k, \\
[\mathsf{J}_i, \mathsf{K}_j] &= \epsilon_{ijk}\, \mathsf{K}_k, \\
[\mathsf{K}_i, \mathsf{K}_j] &= -\epsilon_{ijk}\, \mathsf{J}_k.
\end{aligned} \tag{36.43}$$

Of course, the first of these equations constitutes the Lie algebra of the rotation group $SO(3)$ [see Eq. (9.24)]. The third of these commutators shows that the commutator of two infinitesimal boosts is an infinitesimal rotation, which as mentioned above is the basis of Thomas precession.

The commutation relations of $-i\mathbf{J}$ and $-i\mathbf{K}$ must be the same as those of $\mathsf{\mathbf{J}}$ and $\mathsf{\mathbf{K}}$, or

$$\begin{aligned}
[J_i, J_j] &= i\epsilon_{ijk}\, J_k, \\
[J_i, K_j] &= i\epsilon_{ijk}\, K_k, \\
[K_i, K_j] &= -i\epsilon_{ijk}\, J_k,
\end{aligned} \tag{36.44}$$

which differ only trivially from Eqs. (36.43). The first of these is of course the standard angular momentum commutation relations [see Eq. (10.16)], and the second shows that the vector of matrices $\mathbf{K}$ transforms under rotations as a vector operator. We must now find explicit forms for matrices $\mathbf{J}$ and $\mathbf{K}$ which satisfy Eqs. (36.44).

In the case of ordinary rotations, we only needed to find matrices $\mathbf{J}$ which satisfied the first of these commutation relations, and we did this initially by noticing that $\mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}$ would work [see Eq. (10.17)]. By exponentiation, this gave us the representation (actually, the double-valued representation) of $SO(3)$ by the 2×2 complex matrices in the group $SU(2)$ [see Eq. (10.18)]. In the same spirit, we now notice that the identifications

$$\mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}, \qquad \mathbf{K} = \frac{i}{2}\boldsymbol{\sigma}, \tag{36.45}$$

completely satisfy all three commutation relations (36.44). The matrices $\mathbf{J}$ and $\mathbf{K}$ are 2×2 complex matrices, and generate a 2×2 complex representation of the proper Lorentz group. Using Eqs. (36.45) in Eqs. (36.42), we obtain an explicit representation of infinitesimal Lorentz transformations. That is, when Λ is given by Eq. (36.42a), then $S(\Lambda)$ is given by

$$S(\Lambda) = 1 - \frac{i\epsilon}{2}(\mathbf{a} + i\mathbf{b}) \cdot \boldsymbol{\sigma} = \begin{cases} 1 - \dfrac{i\theta}{2}\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}, & \text{pure rotations,} \\ 1 + \dfrac{\lambda}{2}\hat{\mathbf{m}} \cdot \boldsymbol{\sigma}, & \text{pure boosts,} \end{cases} \tag{36.46}$$

where we have used Eq. (36.35) when specializing to the cases of pure rotations and pure boosts.

Now let us build up finite transformations from infinitesimal ones. First we work with pure rotations, starting with the classical Lorentz transformations $\Lambda = \mathsf{R}(\hat{\mathbf{n}}, \theta)$. If the angle θ is not small, then we cannot use the small angle approximation (36.33a) directly. But we can take a large angle θ and break it up into a large number of small pieces θ/N, where N is large, so that

$$\mathsf{R}(\hat{\mathbf{n}}, \theta) = \left[\mathsf{R}\left(\hat{\mathbf{n}}, \frac{\theta}{N}\right)\right]^N, \tag{36.47}$$

to which Eq. (36.33a) can be applied if N is large enough. Then we obtain

$$\mathsf{R}(\hat{\mathbf{n}}, \theta) = \lim_{N\to\infty}\left[\mathsf{I} + \frac{\theta}{N}\hat{\mathbf{n}} \cdot \mathsf{J}\right]^N. \tag{36.48}$$

However, from elementary calculus we have the limit,

$$\lim_{N\to\infty}\left(1 + \frac{x}{N}\right)^N = e^x, \tag{36.49}$$

which also applies when x is a matrix. Therefore Eq. (36.48) becomes

$$\mathsf{R}(\hat{\mathbf{n}}, \theta) = \exp(\theta\hat{\mathbf{n}} \cdot \mathsf{J}), \tag{36.50}$$

which of course is not new, since it is just the exponential representation of rotations (36.27a) all over again.

The same idea can be applied to the representation $S(\Lambda)$. In the case of pure rotations, where $\Lambda = \mathsf{R}(\hat{\mathbf{n}},\theta)$, let us write $S(\Lambda) = U(\hat{\mathbf{n}},\theta)$. Then we have

$$U(\hat{\mathbf{n}},\theta) = \Big[U\Big(\hat{\mathbf{n}}, \frac{\theta}{N}\Big)\Big]^N = \lim_{N\to\infty}\Big[1 - \frac{i\theta}{2N}\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\Big]^N = e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}, \tag{36.51}$$

where we use Eq. (36.46) for the infinitesimal spinor rotation. In this way we rediscover Eq. (10.18); the pure rotations are represented in the present theory by the usual 2×2 unitary rotation matrices $U(\hat{\mathbf{n}},\theta)$, which belong to the group $SU(2)$.

Now we carry out the same procedure for the boosts. First, for a classical boost of finite (possibly large) rapidity λ, we have

$$\mathsf{B}(\hat{\mathbf{m}},\lambda) = \Big[\mathsf{B}\Big(\hat{\mathbf{m}}, \frac{\lambda}{N}\Big)\Big]^N = \lim_{N\to\infty}\Big[1 + \frac{\lambda}{N}\hat{\mathbf{m}}\cdot\mathbf{K}\Big]^N = \exp(\lambda\hat{\mathbf{m}}\cdot\mathbf{K}), \tag{36.52}$$

which of course just reproduces Eq. (36.27b). As for the spinor representation, when Λ is the pure boost $\mathsf{B}(\hat{\mathbf{m}},\lambda)$, let us write $S(\Lambda) = V(\hat{\mathbf{m}},\lambda)$. Then we have

$$V(\hat{\mathbf{m}},\lambda) = \Big[V\Big(\hat{\mathbf{m}}, \frac{\lambda}{N}\Big)\Big]^N = \lim_{N\to\infty}\Big[1 + \frac{\lambda}{2N}\hat{\mathbf{m}}\cdot\boldsymbol{\sigma}\Big]^N = e^{\lambda\hat{\mathbf{m}}\cdot\boldsymbol{\sigma}/2}, \tag{36.53}$$

where we use Eq. (36.46) for the infinitesimal spinor boost. In this way we find the spinor representation of finite boosts in exponential form.

Let us summarize these results for pure rotations and pure boosts. We have

$$U(\hat{\mathbf{n}},\theta) = e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2} = \cos\frac{\theta}{2} - i(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\sin\frac{\theta}{2}, \tag{36.54a}$$

$$V(\hat{\mathbf{m}},\lambda) = e^{\lambda\hat{\mathbf{m}}\cdot\boldsymbol{\sigma}/2} = \cosh\frac{\lambda}{2} + (\hat{\mathbf{m}}\cdot\boldsymbol{\sigma})\sinh\frac{\lambda}{2}, \tag{36.54b}$$

where we have expanded out the exponentials in the final expressions. Of course, Eq. (36.54a) is the same as Eq. (10.18), but Eq. (36.54b) is new. Equations (36.54) give the spinor representations for pure rotations and pure boosts. Since an arbitrary Lorentz transformation can be represented uniquely as a product of a rotation times a boost, the spinor representation of an arbitrary Lorentz transformation is the product of some U matrix times some V matrix.

We have now found a representation $S(\Lambda)$ of the Lorentz transformations by means of 2×2 complex matrices. Let us explore the properties of this representation. First, we notice that the representation is not unitary, for while the rotation matrices $U(\hat{\mathbf{n}},\theta)$ are unitary, the boost matrices $V(\hat{\mathbf{m}},\lambda)$ are not (they are Hermitian). An arbitrary Lorentz transformation is represented by the product of some U matrix times some V matrix which in general neither unitary nor Hermitian. Thus, $S(\Lambda)^{-1} \neq S(\Lambda)^\dagger$, in general. You may wonder why rotations are represented by unitary matrices, but not Lorentz transformations. In the case of rotations, we argued that the operators representing them on ket spaces should be unitary, in order to preserve probabilities under rotations. Probabilities are also preserved under Lorentz transformations, as they must be, because the outcome of an experiment cannot depend on which Lorentz frame we use to describe the apparatus. However, in order to see the unitarity of a Lorentz transformation, it is necessary to include both the spin and spatial parts of the transformation. We will see this more explicitly when we examine how the Dirac equation transforms under Lorentz transformations. The purely spin part of a Lorentz transformation is, as we see, sometimes represented by non-unitary matrices.

Next, we see that $S(\Lambda)$ is a double-valued representation of the Lorentz group, because it is already double-valued in the rotations alone. That is, angles θ and $\theta + 2\pi$ correspond to the same classical rotation $\mathsf{R}(\hat{\mathbf{n}}, \theta)$, but to two spinor rotations $U(\hat{\mathbf{n}}, \theta + 2\pi) = -U(\hat{\mathbf{n}}, \theta)$ [see Eq. (10.30)]. Thus, for a given Λ, there are actually two 2×2 matrices representing it, which differ by a sign; we may write these as $\pm S(\Lambda)$. The boosts, however, do not introduce any extra multivaluedness. It is really poor notation to write $S(\Lambda)$ unless some indication is given as to which of the two matrices is meant; this situation is exactly the same as with rotations of spin-$\frac{1}{2}$ particles [see the discussion surrounding Eq. (10.31)].

Another property of the matrices $S(\Lambda)$ is that they have unit determinant,

$$\det S(\Lambda) = +1. \tag{36.55}$$

One way to see this is to note that these matrices are generated by the Pauli matrices [albeit with complex coefficients; see Eq. (36.46)]. Since the Pauli matrices are traceless, and since the exponential of a traceless matrix has unit determinant, the matrices $S(\Lambda)$ have unit determinant. Alternatively, we can argue that an arbitrary $S(\Lambda)$ is the product of a U matrix times a V matrix, and both such matrices have unit determinant, as can be verified directly from the expressions in Eq. (36.54). Are there any other constraints besides (36.55) which the matrices $S(\Lambda)$ satisfy? The answer is no. That is, every 2×2 complex matrix with unit determinant can be realized as one of the two $S(\Lambda)$'s for some Lorentz transformation Λ, and, conversely, every Lorentz transformation Λ is represented by some complex, 2×2 matrix with unit determinant. The set of such matrices forms a group, denoted by the rather ungainly notation $SL(2,\mathbb{C})$ (which simply means the special linear group of 2×2 complex matrices; the "special" means unit determinant). The group $SL(2,\mathbb{C})$ is the group of spinor Lorentz transformations. This group has $SU(2)$ as a subgroup, which obviously is the subgroup of pure rotations. This completes our development of the representation $S(\Lambda)$ of the Lorentz group.

It turns out that there is another representation of the Lorentz group by 2×2 matrices, which we will denote by $\tilde{S}(\Lambda)$ to distinguish it from $S(\Lambda)$. The representations $S(\Lambda)$ and $\tilde{S}(\Lambda)$ are inequivalent (not simply related by a change of basis). Both representations are needed to transform Dirac spinors. To obtain the new representation, we adopt a different representation of the Lie algebra (36.44); namely, instead of Eq. (36.45), we take

$$\mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}, \qquad \mathbf{K} = -\frac{i}{2}\boldsymbol{\sigma}, \tag{36.56}$$

which, like Eq. (36.45), also causes the commutation relations (36.44) to be satisfied. This new representation of the Lie algebra differs from the old one simply by a change in sign in the $\mathbf{K}$ matrices, which of course generate the boosts. With the new identifications of the matrices $\mathbf{J}$ and $\mathbf{K}$, the representation of an infinitesimal Lorentz transformation is

$$\tilde{S}(\Lambda) = 1 - \frac{i\epsilon}{2}(\mathbf{a} - i\mathbf{b})\cdot\boldsymbol{\sigma}, \tag{36.57}$$

instead of Eq. (36.46). By building up finite transformations from infinitesimal ones, exactly as we did above, we can find the matrices $\tilde{S}(\Lambda)$ for finite Lorentz transformations. Calling these matrices $\tilde{U}(\hat{\mathbf{n}},\theta)$ and $\tilde{V}(\hat{\mathbf{m}},\lambda)$ in the case of pure rotations and boosts, respectively, we find

$$\tilde{U}(\hat{\mathbf{n}},\theta) = e^{-i\theta\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2} = \cos\frac{\theta}{2} - i(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma})\sin\frac{\theta}{2} = U(\hat{\mathbf{n}},\theta), \tag{36.58a}$$

$$\tilde{V}(\hat{\mathbf{m}},\lambda) = e^{-\lambda\hat{\mathbf{m}}\cdot\boldsymbol{\sigma}/2} = \cosh\frac{\lambda}{2} - (\hat{\mathbf{m}}\cdot\boldsymbol{\sigma})\sinh\frac{\lambda}{2} = V(\hat{\mathbf{m}},\lambda)^{-1}, \tag{36.58b}$$

instead of Eqs. (36.54). We see that the two representations $S(\Lambda)$ and $\tilde{S}(\Lambda)$ are identical in the case of pure rotations, but that the boosts are represented by inverse matrices in the two cases.

A general Lorentz transformation is represented by a product of a U-matrix times a V-matrix, either in the form $S = UV$ or $\tilde{S} = \tilde{U}\tilde{V}$. It turns out that in the general case, the relation between the two representations is given by

$$\tilde{S}(\Lambda) = S(\Lambda)^{\dagger -1}. \tag{36.59}$$

This relation is easily proven in the case of pure rotations, for which $U^{-1} = U^\dagger$ and $\tilde{U} = U$; and in the case of pure boosts, for which $V^\dagger = V$ and $\tilde{V} = V^{-1}$. To prove it in the case of a general Lorentz transformation, we can show that if Eq. (36.59) is true for two Lorentz transformations Λ_1 and Λ_2, then it is true for $\Lambda_1\Lambda_2$. This proof runs as follows:

$$\begin{aligned}\tilde S(\Lambda_1)\tilde S(\Lambda_2) &= S(\Lambda_1^{-1})^\dagger S(\Lambda_2^{-1})^\dagger = [S(\Lambda_2^{-1})S(\Lambda_1^{-1})]^\dagger \\ &= S(\Lambda_2^{-1}\Lambda_1^{-1})^\dagger = S\bigl((\Lambda_1\Lambda_2)^{-1}\bigr)^\dagger = \tilde S(\Lambda_1\Lambda_2).\end{aligned} \tag{36.60}$$

We now turn to the subject of adjoint formulas, which are important in the case of ordinary rotations because they indicate that the angular momentum is actually a vector operator. In the case of spin-$\frac{1}{2}$ systems, the adjoint formula for rotations is given by Eq. (10.24), which we will now generalize to Lorentz transformations. There are actually two generalizations. First we write $\sigma_0 = \sigma^0 = 1$ (the 2×2 identity matrix), and define a 4-vector of Pauli matrices by

$$\sigma_\mu = (1, \boldsymbol{\sigma}), \qquad \sigma^\mu = (1, -\boldsymbol{\sigma}), \tag{36.61}$$

where, as indicated, we raise and lower indices in the usual way. Similarly, we let $\tau_0 = \tau^0 = 1$, and we define a new 4-vector of Pauli matrices by

$$\tau_\mu = (1, -\boldsymbol{\sigma}), \qquad \tau^\mu = (1, \boldsymbol{\sigma}). \tag{36.62}$$

The notation suggests that these two vectors of matrices transform as 4-vectors. The sense in which this is true is given by the two adjoint formulas,

$$S(\Lambda)\,\sigma^\mu\, S(\Lambda)^\dagger = (\Lambda^{-1})^\mu{}_\nu\,\sigma^\nu, \tag{36.63a}$$

$$\tilde S(\Lambda)\,\tau^\mu\,\tilde S(\Lambda)^\dagger = (\Lambda^{-1})^\mu{}_\nu\,\tau^\nu, \tag{36.63b}$$

which should be compared to Eq. (10.24).

We will prove only Eq. (36.63a). We let $X^\mu = (X^0, \mathbf{X})$ and $X_\mu = (X^0, -\mathbf{X})$ be an arbitrary 4-vector, and we consider at first only infinitesimal Lorentz transformations. Then we have

$$\begin{aligned}S(\Lambda)\,X_\mu\sigma^\mu\,S(\Lambda)^\dagger &= \Bigl[1 - \frac{i\epsilon}{2}(\mathbf{a}+i\mathbf{b})\cdot\boldsymbol{\sigma}\Bigr]X_\mu\sigma^\mu\Bigl[1 + \frac{i\epsilon}{2}(\mathbf{a}-i\mathbf{b})\cdot\boldsymbol{\sigma}\Bigr] \\ &= X_\mu\sigma^\mu + \epsilon\Bigl(-\frac{i}{2}\bigl[\mathbf{a}\cdot\boldsymbol{\sigma}, X^0 + \mathbf{X}\cdot\boldsymbol{\sigma}\bigr] + \frac{1}{2}\bigl\{\mathbf{b}\cdot\boldsymbol{\sigma}, X^0 + \mathbf{X}\cdot\boldsymbol{\sigma}\bigr\}\Bigr),\end{aligned} \tag{36.64}$$

where X^0 means X^0 times the identity matrix and where the square and curly brackets in the final expression are respectively the matrix commutator and anticommutator. But by Eqs. (36.28) we have

$$-\frac{i}{2}\bigl[\mathbf{a}\cdot\boldsymbol{\sigma}, X^0 + \mathbf{X}\cdot\boldsymbol{\sigma}\bigr] = (\mathbf{a}\times\mathbf{X})\cdot\boldsymbol{\sigma} = -(\mathbf{a}\cdot\mathsf{J})^\mu{}_\nu\,X_\mu\sigma^\nu, \tag{36.65}$$

and by Eq. (36.29) we have

$$\frac{1}{2}\{\mathbf{b}\cdot\boldsymbol{\sigma}, X^0+\mathbf{X}\cdot\boldsymbol{\sigma}\} = X^0\,(\mathbf{b}\cdot\boldsymbol{\sigma}) + \mathbf{b}\cdot\mathbf{X} = -(\mathbf{b}\cdot\mathsf{K})^{\mu}{}_{\nu}\,X_{\mu}\sigma^{\nu}. \tag{36.66}$$

Therefore Eq. (36.64) becomes

$$S(\Lambda)\,X_{\mu}\sigma^{\mu}\,S(\Lambda)^{\dagger} = X_{\mu}\big[\mathsf{I}-\epsilon(\mathbf{a}\cdot\mathsf{J}+\mathbf{b}\cdot\mathsf{K})\big]^{\mu}{}_{\nu}\,\sigma_{\nu} = X_{\mu}\,(\Lambda^{-1})^{\mu}{}_{\nu}\,\sigma^{\nu}, \tag{36.67}$$

which proves Eq. (36.63a) for infinitesimal Lorentz transformations. But it is easy to show that if Eq. (36.63a) is true for two Lorentz transformations, then it is true for their product; therefore, by building up finite transformations from infinitesimal ones, Eq. (36.63a) is true for finite transformations as well. The proof of Eq. (36.63b) is similar.

For completeness, we give some further useful relations satisfied by the spinor representations $S(\Lambda)$ and $\tilde{S}(\Lambda)$ which were not discussed in class. We begin by noticing that from the definitions (36.61) and (36.62), we have

$$\mathrm{tr}(\sigma^{\mu}\tau_{\nu}) = \delta^{\mu}_{\nu}. \tag{36.68}$$

This allows us to multiply Eq. (36.63a) by τ_{α} and take traces to solve for the classical matrix Λ in terms of the spinor representation $S(\Lambda)$. We can do something similar with Eq. (36.63b). Upon rearranging indices, the result can be placed in the following, convenient forms:

$$(\Lambda^{-1})^{\mu}{}_{\nu} = \frac{1}{2}\,\mathrm{tr}\big[S(\Lambda)\sigma^{\mu}S(\Lambda)^{\dagger}\tau_{\nu}\big], \tag{36.69a}$$

$$= \frac{1}{2}\,\mathrm{tr}\big[\tilde{S}(\Lambda)\tau^{\mu}\tilde{S}(\Lambda)^{\dagger}\sigma_{\nu}\big], \tag{36.69b}$$

which are generalizations of Eq. (10.34) in the nonrelativistic theory. These equations can also be written in another form,

$$\Lambda^{\mu}{}_{\nu} = \frac{1}{2}\,\mathrm{tr}\big[\tau^{\mu}S(\Lambda)\sigma_{\nu}S(\Lambda)^{\dagger}\big], \tag{36.70a}$$

$$= \frac{1}{2}\,\mathrm{tr}\big[\sigma^{\mu}\tilde{S}(\Lambda)\tau_{\nu}\tilde{S}(\Lambda)^{\dagger}\big]. \tag{36.70b}$$

This completes our introduction to the spinor representations of the Lorentz group.

Now we turn to the covariance of the Dirac equation under Lorentz transformations. At first we deal with the free particle Dirac equation only, which we write in the form

$$i\hbar\frac{\partial\psi}{\partial t} = -i\hbar c\,\boldsymbol{\alpha}\cdot\nabla\psi + mc^2\beta\psi, \tag{36.71}$$

where $\boldsymbol{\alpha}$ and β are the usual, Hermitian, 4×4 Dirac matrices which satisfy the anticommutation relations,

$$\{\alpha_k, \alpha_\ell\} = 2\delta_{k\ell}, \qquad \{\alpha_k, \beta\} = 0, \qquad \{\beta, \beta\} = 2. \tag{36.72}$$

As discussed in class, these anticommutation relations plus the requirement that $\boldsymbol{\alpha}$ and β be Hermitian determine the matrices $\boldsymbol{\alpha}$ and β up to a unitary change of basis (all 4×4 Hermitian representations of the anticommutation relations are unitarily equivalent).

Next we introduce the definitions,

$$\gamma^0 = \beta, \qquad \gamma^i = \beta\alpha_i, \tag{36.73}$$

and arrange the new γ matrices as a 4-vector,

$$\gamma^\mu = (\gamma^0, \boldsymbol{\gamma}). \tag{36.74}$$

Notice that γ^μ is a 4-vector of 4×4 Dirac matrices. The notation suggests that γ^μ transforms as a 4-vector under Lorentz transformations; the sense in which this is true remains to be shown. We notice that $\gamma^0 = \beta$ is Hermitian, but the spatial components γ^i are anti-Hermitian, since we have

$$(\gamma^i)^\dagger = (\beta\alpha_i)^\dagger = \alpha_i\beta = -\beta\alpha_i = -\gamma^i, \tag{36.75}$$

where we use the anticommutation relation $\{\alpha_i, \beta\} = 0$. The matrices γ^μ satisfy a set of anticommutation relations which are equivalent to (36.72), but which can be stated more compactly,

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}. \tag{36.76}$$

It is possible (and in some ways desirable) to develop the whole theory of the Dirac matrices by starting with the anticommutation relations (36.76). These algebraic relations encode the relativistic energy-momentum relations, and can be taken as the defining property of the γ^μ matrices. These relations essentially define the γ^μ matrices, because all 4-vectors of matrices which satisfy these relations are merely related by a change of basis, which is unitary if we assume (as we will) that γ^0 is Hermitian. All physical results in the Dirac theory follow from these algebraic relations alone, and in fact it is never necessary to descend to specific representations. (The same is true of the nonrelativistic Pauli matrices; there are no physical results which depend on the usual standard forms of these matrices. Instead, all the physics follows from the algebraic relations satisfied by these matrices.) Nevertheless, sometimes it is desirable to have specific representations for the γ^μ matrices. In the Dirac-Pauli representation, we have

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{36.77}$$

and in the Weyl representation we have

$$\gamma^0 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}. \tag{36.78}$$

These are the two most popular representations in the literature; sometimes a third, the Majorana representation, is also used. All these matrices are 4×4 matrices partitioned into four 2×2 blocks.

The form (36.71) of the Dirac equation is regarded as the Hamiltonian or Hermitian form, since the right hand side has the form $H\psi$, where H is the Hermitian Dirac Hamiltonian. The Hamiltonian or Hermitian form is in contrast to the covariant form, which we obtain by multiplying Eq. (36.71) from the left by $\gamma^0 = \beta$. The result can be written as

$$i\hbar\gamma^\mu \frac{\partial\psi}{\partial x^\mu} - mc\,\psi = 0. \tag{36.79}$$

By using the 4-vector of operators $p_\mu = i\hbar\partial/\partial x^\mu$, this can also be written in the form

$$(\gamma^\mu p_\mu - mc)\psi = 0. \tag{36.80}$$

The notation is suggestive, and we are calling this the covariant form of the Dirac equation, but notice that as yet we have not proven covariance.

A further change of notation is often convenient. If X^μ is a 4-vector (of numbers, operators, etc.), then we denote its contraction with the vector γ^μ of Dirac matrices with a slash, so that

$$\not{X} = \gamma^\mu X_\mu. \tag{36.81}$$

This is called the *Feynman slash*. When you see this notation, you should remember that the object is actually a Dirac 4×4 matrix, with components determined by the 4-vector X^μ in question.

In terms of the Feynman slash, we can write the free particle Dirac equation in a very compact form,

$$(\not{p} - mc)\psi = 0. \tag{36.82}$$

If we wish to include the coupling with an electromagnetic field, we can use the minimal coupling prescription,

$$p^\mu \to p^\mu - \frac{q}{c}A^\mu, \tag{36.83}$$

where $A^\mu = (\Phi, \mathbf{A})$ is the electromagnetic 4-vector potential, which causes the Dirac equation to become

$$\Big(\not{p} - \frac{q}{c}\not{A} - mc\Big)\psi = 0. \tag{36.84}$$

This equation is a good deal simpler in appearance than its nonrelativistic counterpart (the Schrödinger-Pauli equation for an electron, including the magnetic interaction with the spin), and yet it contains more physics than the latter (for example, all the fine structure corrections).

Now let us consider how Dirac spinors transform under Lorentz transformations. We begin by reminding ourselves how nonrelativistic Pauli spinors transform under ordinary rotations. We let $|\psi\rangle$ be the state of a nonrelativistic electron, and we transform it under a rotation specified by R to a new state $|\psi'\rangle$,

$$|\psi'\rangle = U(\mathsf{R})|\psi\rangle. \tag{36.85}$$

We denote the 2-component wave function corresponding to $|\psi\rangle$ by $\psi_m(\mathbf{r})$ in the $(\mathbf{r}, S_z)$-representation, so that

$$\psi_m(\mathbf{r}) = \langle \mathbf{r}, m|\psi\rangle. \tag{36.86}$$

Then the relation between the old and new wave functions is

$$\psi'_m(\mathbf{r}) = \sum_{m_1} D^s_{mm_1}(\mathsf{R})\psi_{m_1}\big(\mathsf{R}^{-1}\mathbf{r}\big), \tag{36.87}$$

where $s = \frac{1}{2}$ and where the D-matrix is one of the 2×2 matrices U introduced in Notes 10.

As for the Dirac equation, let the wave function be $\psi(x)$, where x stands for x^μ, which includes both the space- and time-dependence. The wave function itself is of course a 4-component spinor. Then the nonrelativistic transformation law (36.87) suggests that the Dirac spinor should transform under Lorentz transformations according to

$$\psi(x) \to \psi'(x) = D(\Lambda)\psi\big(\Lambda^{-1}x\big), \tag{36.88}$$

where $D(\Lambda)$ is some 4×4 representation of the Lorentz group. It is understood that the matrix D is multiplied times the spinor ψ in this equation.

Let us check that this equation makes sense, at least in its space-time dependence. Suppose we have a plane-wave, free-particle solution in the form,

$$\psi \sim e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}, \tag{36.89}$$

where the $\sim$ means that we will concentrate on the space-time dependence and ignore the spinor part as much as possible. If we introduce the wave 4-vector k^μ,

$$k^\mu = \Big(\frac{\omega}{c}, \mathbf{k}\Big), \tag{36.90}$$

then the phase of ψ can be written in covariant form,

$$k_\mu x^\mu = \omega t - \mathbf{k}\cdot\mathbf{x}. \tag{36.91}$$

(Notice that with this notation, the Einstein-de Broglie relations become a single equation, $p^\mu = \hbar k^\mu$.) Therefore under the transformation (36.88), the x-dependence of the phase transforms by Λ^{-1}, or,

$$\psi \sim \exp\bigl(-ik_\mu x^\mu\bigr) \to \psi' \sim \exp\bigl[-ik_\mu (\Lambda^{-1})^\mu{}_\nu x^\nu\bigr]. \tag{36.92}$$

But we can use Eq. (36.15) to express Λ^{-1} in terms of Λ, which allows us to write the phase in the form,

$$k_\mu (\Lambda^{-1})^\mu{}_\nu x^\nu = k'_\mu x^\mu, \tag{36.93}$$

where

$$k'^\mu = \Lambda^\mu{}_\nu\, k^\nu. \tag{36.94}$$

Therefore since Λ represents a rotation or a boost in an active sense, we see that the transformed Dirac wave function ψ' is related to the original wave function ψ by having its momentum rotated or boosted in an active sense. This is what we expect, and it shows that the Λ^{-1} in Eq. (36.88) makes sense.

Now that our confidence is enhanced that Eq. (36.88) is the correct transformation law for Dirac wave functions under Lorentz transformations [for some as yet unknown, 4×4 representation $D(\Lambda)$ of the Lorentz group], let us make use of the Dirac equation to see what conditions are imposed on the matrices $D(\Lambda)$. Specifically, let us suppose that $\psi(x)$ is a solution of the free particle Dirac equation,

$$i\hbar\gamma^\mu \frac{\partial\psi(x)}{\partial x^\mu} - mc\,\psi(x) = 0, \tag{36.95}$$

and let us demand that the Lorentz transformed wave function $\psi'(x)$, given by Eq. (36.88), also be a solution of the free particle Dirac equation. [If Λ were interpreted in a passive instead of an active sense, this would be equivalent to demanding that the Dirac equation have the same form in all Lorentz frames.] Then we have

$$i\hbar\gamma^\mu D(\Lambda)\frac{\partial\psi\bigl(\Lambda^{-1}x\bigr)}{\partial x^\mu} - mc\,D(\Lambda)\psi\bigl(\Lambda^{-1}x\bigr) = 0. \tag{36.96}$$

We multiply this through by $D(\Lambda)^{-1}$ and set $y = \Lambda^{-1}x$, or $y^\nu = (\Lambda^{-1})^\nu{}_\mu x^\mu$, so that

$$i\hbar D(\Lambda)^{-1}\gamma^\mu D(\Lambda)\,\bigl(\Lambda^{-1}\bigr)^\nu{}_\mu \frac{\partial\psi(y)}{\partial y^\nu} - mc\,\psi(y) = 0. \tag{36.97}$$

But x in Eq. (36.95) is a dummy variable which we can replace by y, so that

$$mc\,\psi(y) = i\hbar\gamma^\nu \frac{\partial\psi(y)}{\partial y^\nu}. \tag{36.98}$$

Then Eq. (36.97) becomes

$$D(\Lambda)^{-1}\gamma^\mu D(\Lambda)\left(\Lambda^{-1}\right)^\nu{}_\mu = \gamma^\nu, \tag{36.99}$$

after cancelling $i\hbar$ and the space-time derivatives $\partial\psi(y)/\partial y^\nu$, which can take on any values. Equation (36.99) is a condition on the representation $D(\Lambda)$ which can be put into a more convenient form by replacing Λ by Λ^{-1} and rearranging terms, whereupon we find

$$D(\Lambda)\gamma^\mu D(\Lambda)^{-1} = (\Lambda^{-1})^\mu{}_\nu\,\gamma^\nu. \tag{36.100}$$

This should be compared to its nonrelativistic counterpart, Eq. (10.24). It is this equation which qualifies γ^μ as a 4-vector; you may recall the definition (15.9) of a vector operator in the nonrelativistic theory. In the present context, Eq. (36.100) serves as a requirement which the as yet unknown representation $D(\Lambda)$ must satisfy, in order that the Dirac equation be covariant under Lorentz transformations.

We must find the 4×4 representation $D(\Lambda)$ which satisfies Eq. (36.100). One approach is simply to work with the Dirac matrices in some representation. The Weyl representation is more convenient for this purpose; we notice that with the definitions (36.61) and (36.62), the Dirac matrices in the Weyl representation (36.78) can be written as

$$\gamma^\mu = \begin{pmatrix} 0 & -\sigma^\mu \\ -\tau^\mu & 0 \end{pmatrix}. \tag{36.101}$$

Then we simply notice that if we take

$$D(\Lambda) = \begin{pmatrix} S(\Lambda) & 0 \\ 0 & \tilde{S}(\Lambda) \end{pmatrix} \tag{36.102}$$

(this is also in the Weyl representation), then by direct matrix multiplication we have

$$D(\Lambda)\gamma^\mu D(\Lambda)^{-1} = \begin{pmatrix} 0 & -S(\Lambda)\sigma^\mu\tilde{S}(\Lambda)^{-1} \\ -\tilde{S}(\Lambda)\tau^\mu S(\Lambda)^{-1} & 0 \end{pmatrix}. \tag{36.103}$$

But by Eqs. (36.59) and (36.63), we have

$$\begin{aligned} S(\Lambda)\sigma^\mu\tilde{S}(\Lambda)^{-1} &= S(\Lambda)\sigma^\mu S(\Lambda)^\dagger = (\Lambda^{-1})^\mu{}_\nu\,\sigma^\nu, \\ \tilde{S}(\Lambda)\tau^\mu S(\Lambda)^{-1} &= \tilde{S}(\Lambda)\tau^\mu\tilde{S}(\Lambda)^\dagger = (\Lambda^{-1})^\mu{}_\nu\,\tau^\mu, \end{aligned} \tag{36.104}$$

so Eq. (36.100) is verified. Thus, Eq. (36.102) gives us an explicit relation between the 4×4 representation $D(\Lambda)$ and the two 2×2 representations $S(\Lambda)$ and $\tilde{S}(\Lambda)$.

Another approach to the determination of the representation $D(\Lambda)$ is that taken by Bjorken and Drell. Their approach is superior to the one given here in that they work only with the fundamental anticommutation relations (36.76), whereas we have worked in a specific representation. Of course, in view of the equivalence of all representations of the Dirac algebra, it is never wrong to work in a specific representation, but one might say that it is more esthetically pleasing to use only the abstract defining properties of the γ^μ matrices. On the other hand, the presentation of Bjorken and Drell is flawed in that they never show that the $D(\Lambda)$ which they construct actually forms a representation of the Lorentz group. To do that, they would have to show that their infinitesimal generators, expressed in terms of the Dirac matrices $\sigma^{\mu\nu}$, actually do satisfy the expected commutation relations [essentially Eqs. (36.44)]. Another flaw of Bjorken and Drell (see pp. 16 and 17) is that they say that the distinction between proper and improper Lorentz transformations is determined solely by $\det \Lambda$ (they ignore $\Lambda^0{}_0$). Otherwise, the discussion in Chapter 2 of Bjorken and Drell is ok, and at this point you should be able to follow it without trouble.