3 The Lorentz Group and the Pauli Algebra

3.1 Introduction

Twentieth century physics is dominated by the development of relativity and quantum mechanics, disciplines centered around the universal constants c and h respectively. Historically, the emergence of these constants revealed a socalled breakdown of classical concepts.

From the point of view of our present knowledge, it would be evidently desirable to avoid such breakdowns and formulate only principles which are correct according to our present knowledge. Unfortunately, no one succeeded thus far to suggest a ”correct” postulational basis which would be complete enough for the wide ranging practical needs of physics.

The purpose of this course is to explore a program in which we forego, or rather postpone, the requirement of completeness, and consider at first only simple situations. These are described in terms of concepts which form the basis for the development of a precise mathematical formalism with empirically verified physical implications. The continued alternation of conceptual analysis with formal developments gradually extends and deepens the range of situations covered, without affecting consistency and empirical validity.

According to the central idea of quantum mechanics all particles have undulatory properties, and electromagnetic radiation has corpuscular aspects. In the quantitative development of this idea we have to make a choice, whether to start with the classical wave concept and build in the corpscular aspects, or else start with the classical concept of the point particle, endowed with a constant and invariant mass, and modify these properties by means of the wave concept. Eventually, the resulting theory should be independent of the path chosen, but the details of the construction process are different.

The first alternative is apparent in Einstein’s photon hypothesis[9], which is closely related with his special theory of relativity[10].

In contrast, the wealth of nonrelativistic problems within atomic, molecular and nuclear physics favored the second approach which is exploited in the Bohr-Heisenberg quantum mechanics.

The course of the present developments is set by the decision of following up the Einsteinian departure.

3.2 The corpuscular aspects of light

As a first step in carrying out the problem just stated, we start with a precise, even though schematic, formulation of wave kinematics.

[9]See Reference[Ein05a].

[10]See reference [Ein05b]. Einstein discussed the implications of relativistic and quantum effects on the theory of radiation in an address before the 81st assembly of the Society of German Scientists and Physicians, Salzburg, 1909 [Ein09].

We consider first a spherical wave front

$$r_o^2 - \vec{r}^2 = 0 \tag{3.2.1}$$

where

$$r_o = ct \tag{3.2.2}$$

and t is the time elapsed since emission to the instantaneous wave front.

In order to describe propagation in a definite direction, say along the unit vector $\hat{k}$, we introduce an appropriately chosen tangent plane corresponding to a monochromatic plane wave

$$k_o r_o - \vec{k} \cdot \vec{r} = 0 \tag{3.2.3}$$

with

$$k_o = \omega/c \tag{3.2.4}$$

$$\vec{k} = \frac{2\pi}{\lambda}\hat{k} \tag{3.2.5}$$

and

$$k_o^2 - \vec{k}^2 = 0 \tag{3.2.6}$$

where the symbols have their conventional meanings.

Next we postulate that radiation has a granular character, as it is expressed already in Definition 1 of Newton's Optics! However, in a more quantitative sense we state the standard quantum condition according to which a quantum of light pulse with the wave vector $(k_o, \vec{k})$ is associated with a four-momentum

$$(p_o, \vec{p}) = \hbar(k_o, \vec{k}) \tag{3.2.7}$$

$$p_o^2 - \vec{p}^2 = 0 \tag{3.2.8}$$

with

$$p_o = \frac{E}{c} \tag{3.2.9}$$

where E is the energy of the light quantum, or photon.

The proper coordination of the two descriptions involving spherical and plane waves, presents problems to which we shall return later. At this point it is sufficient to note that individual photons have directional properties described by a wave vector,and a spherical wave can be considered as an assembly of photons emitted isotropically from a small source.

As the next step in our procedure we argue that the photon as a particle should be associated with an object group, as introduced in Sec. 1.7. Assuming with Einstein that light velocity is unaffected by an inertial transformation, the passive kinematic group that leaves Eq's (1) - (3) invariant is the Lorentz group.

There are few if any principles in physics which are as thoroughly justified by their implications as the principle of Lorentz invariance. Our objective is to develop these implications in a systematic fashion.

In the early days of relativity the consequences of Lorentz invariance involved mostly effects of the order of (v/c)2 , a quantity that is small for the velocities attainable at that time. The justification is much more dramatic at present when we can refer to the operation of high energy accelerators operating near light velocity.

Yet this is not all. Lorentz invariance has many consequences which are valid even in nonrelativistic physics, but classically they require a number of independent postulates for their justification. In such cases the Lorentz group is used to achieve conceptual economy.

In-view of this far-reaching a posteriori verification of the constancy of light velocity, we need not be unduly concerned with its a priori justification. It is nevertheless a puzzling question, and it has given rise to much speculation: what was Einstein's motivation in advancing this postulate? [11]

Einstein himself gives the following account of a paradox on which he hit at the age of sixteen[12]:

> "If I pursue a beam of light with the velocity c (velocity of light in vacuum), I should observe such a beam of light' as a spatially oscillatory electromagnetic field at rest. However, there seems to be no such thing, whether on the basis of experience or according to Maxwell's equations."

The statement could be actually even sharpened: on overtaking a travelling wave, the resulting phenomenon would simply come to rest, rather than turn into a standing wave.

However it may be, if the velocity of propagation were at all affected by the motion of the observer, it could be "transformed away." Should we accept such a radical change from an inertial transformation? At least in hindsight, we know that the answer is indeed no.

Note that the essential point in the above argument is that a light quantum cannot be transformed to rest. This absence of a preferred rest system with respect to the photon does not exclude the existence of a preferred frame defined from other considerations. Thus it has been recently suggested that a preferred frame be defined by the requirement that the $3K$ radiation be isotropic in it [Wei72].

Since Einstein and his contemporaries emphasized the absence of any preferred frame of reference, one might have wondered whether the aforementioned radiation, or some other cosmologically defined frame,might cause difficulties in the theory of relativity.

Our formulation, based on weaker assumptions, shows that such concern is unwarranted.

[11] *See Gerald Holton, Einstein, Michelson and "Crucial" Experiment, in Thematic Origins of Scientific Thought. Kepler to Einstein. Harvard University Press, Cambridge, Mass., 1973, pp 261-352.

[12]

We observe finally, that we have considered thus far primarily wave kinematics, with no reference to the electrodynamic interpretation of light. This is only a tactical move. We propose to derive classical electrodynamics (CED) within our scheme, rather than suppose its validity.

Problems of angular momentum and polarization are also left for later inclusion.

However, we are ready to widen our context by being more explicit with respect to the properties of the four-momentum.

Eq (5) provides us with a definition of the four-momentum, but only for the case of the photon, that is for a particle with zero rest mass and the velocity c.

This relation is easily generalized to massive particles that can be brought to rest. We make use of the fact that the Lorentz transformation leaves the left-hand side of Eq (5) invariant, whether or not it vanishes. Therefore we set

$$p_o^2 - \vec{p}\cdot\vec{p} = m^2c^2 \tag{3.2.10}$$

and define the mass m of a particle as the invariant "length" of the four-momentum according to the Minkowski metric (with $c = 1$).

We can now formulate the postulate: **The four-momentum is conserved**. This principle includes the conservation of energy and that of the three momentum components. It is to be applied to the interaction between the photon and a massive particle and also to collision processes in general.

In order to make use of the conservation law, we need explicit expressions for the velocity dependence of the four-momentum components. These shall be obtained from the study of the Lorentz group.

3.3 On circular and hyperbolic rotations

We propose to develop a unified formalism for dealing with the Lorentz group $\mathcal{SO}(3,1)$ and its subgroup $\mathcal{SO}(3)$. This program can be divided into two stages. First, consider a Lorentz transformation as a hyperbolic rotation, and exploit the analogies between circular and hyperbolic trigonometric functions, and also of the corresponding exponentials. This simple idea is developed in this section in terms of the subgroups $\mathcal{SO}(2)$ and $\mathcal{SO}(1,1)$. The rest of this chapter is devoted to the generalization of these results to three spatial dimensions in terms of a matrix formalism.

Let us consider a two-component vector in the Euclidean plane:

$$\vec{x} = x_1\hat{e}_1 + x_2\hat{e}_2 \tag{3.3.1}$$

We are interested in the transformations that leave $x_1^2 + x_2^2$ invariant. Let us write

$$x_1^2 + x_2^2 = (x_1 + ix_2)(x_1 - ix_2) \tag{3.3.2}$$

and set

$$(x_1 + ix_2)' = a(x_1 + ix_2) \tag{3.3.3}$$
$$(x_1 - ix_2)' = a^*(x_1 - ix_2) \tag{3.3.4}$$

where the star means conjugate complex. For invariance we have

$$aa^* = 1 \tag{3.3.5}$$

or

$$a = \exp(-i\phi)\,, \quad a^* = \exp(i\phi) \tag{3.3.6}$$

From these formulas we easily recover the elementary trigonometric expressions. Table 3.1 summarizes the presentations of rotational transformations in terms of exponentials, trigonometric functions and algebraic irrationalities involving the slope of the axes. There is little to recommend the use of the latter, however it completes the parallel with Lorentz transformations where this parametrization is favored by tradition.

We emphasize the advantages of the exponential function, mainly because it lends itself to iteration, which is apparent from the well known formula of de Moivre:

$$\exp(in\phi) = \cos(n\phi) + i\sin(n\phi) = (\cos(\phi) + i\sin(\phi))^n \tag{3.3.7}$$

The same Table contains also the parametrization of the Lorentz group in one spatial variable. The analogy between $\mathcal{SO}(2)$ and $\mathcal{SO}(1,1)$ is far reaching and the Table is selfexplanatory. Yet there are a number of additional points which are worth making.

The invariance of

$$x_0^2 - x_3^2 = (x_0 + x_3)\,(x_0 - x_3) \tag{3.3.8}$$

is ensured by

$$(x_0 + x_3)' = a(x_0 + x_3) \tag{3.3.9}$$
$$(x_0 - x_3)' = a^{-1}(x_0 - x_3) \tag{3.3.10}$$

for an arbitrary a. By setting $a = \exp(-\mu)$ in the Table we tacitly exclude negative values. Admitting a negative value for this parameter would imply the interchange of future and past. The Lorentz transformations which leave the direction of time invariant, are called *orthochronic*. Until further notice these are the only ones we shall consider.

The meaning of the parameter μ is apparent from the well known relation

$$\tanh\mu = \frac{v}{c} = \beta \tag{3.3.11}$$

where v is the velocity of the primed system Σ' measured in Σ. Being a (non-Euclidean) measure of a velocity, μ is sometimes called **rapidity**, or **velocity parameter**.

Rotation	Lorentz Transformation
$(x_1 - ix_2)' = e^{i\phi}(x_1 - ix_2)$	$(x_0 + x_3)' = e^{-\mu}(x_0 + x_3)$
$(x_1 + ix_2)' = e^{-i\phi}(x_1 + ix_2)$	$(x_0 - x_3)' = e^{\mu}(x_0 - x_3)$
$x_1' = x_1 \cos\phi + x_2 \sin\phi$	$x_3' = x_3 \cosh\mu - x_0 \sinh\mu$
$x_2' = -x_1 \sin\phi + x_2 \cos\phi$	$x_0' = -x_3 \sinh\mu + x_0 \cosh\mu$
$x_1' = \frac{x_1+\kappa x_2}{\sqrt{1+\kappa^2}}$	$x_3' = \frac{x_3-\beta x_0}{\sqrt{1-\beta^2}}$
$x_2' = \frac{\kappa x_1+x_2}{\sqrt{1+\kappa^2}}$	$x_0' = \frac{-\beta x_3+x_0}{\sqrt{1-\beta^2}}$
$\kappa = \tan\phi = \left(\frac{x_2}{x_1}\right)_{x_2' = 0}$	$\beta = \tanh\mu = \left(\frac{x_3}{x_0}\right)_{x_3' = 0}$
$= -\left(\frac{x_2'}{x_1'}\right)_{x_2 = 0}$	$= \left(\frac{x_3'}{x_0'}\right)_{x_3 = 0}, \quad x_0 = ct$
$\frac{X_2}{x_1} = \tan\theta$	$\frac{X_3}{x_0} = \tanh\nu$
$\theta' = \theta - \phi$	$\nu' = \nu - \mu$

Table 3.1: Summary of the rotational transformations. (The signs of the angles correspond to the passive interpretation.)

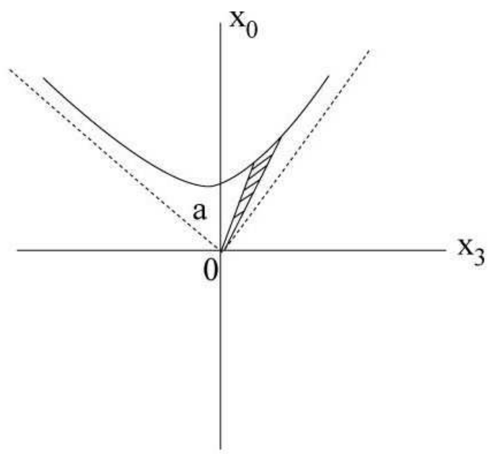

Figure 3.1: Area in (x_0, x_3)-plane.

We shall refer to μ also as **hyperbolic angle**. The formal analogy with the **circular angle** ϕ is evident from the Table. We deepen this parallel by means of the observation that μ can be interpreted as an area in the (x_0, x_3) plane (see Figure 3.1).

Consider a hyperbola with the equation

$$\left(\frac{x_0}{a}\right)^2 - \left(\frac{x_3}{b}\right)^2 = 1 \tag{3.3.12}$$

$$x_0 = a\cosh\mu \qquad x_3 = b\sinh\mu \tag{3.3.13}$$

The shaded triangular area (shown in Figure 3.1) is according to Equation 2.6.2 of Section 2.6:

$$\frac{1}{2}\begin{vmatrix} x_3 + dx_3 & x_3 \\ x_0 + dx_0 & x_0 \end{vmatrix} = \frac{1}{2}\left(x_0 dx_3 - x_3 dx_0\right) = \tag{3.3.14}$$

$$\frac{ab}{2}\left(\cosh\mu^2 - \sinh\mu^2\right) d = \frac{ab}{2} d\mu \tag{3.3.15}$$

We could proceed similarly for the circular angle ϕ and define it in terms of the area of a circular sector, rather than an arc. However, only the area can be generalized for the hyperbola.

Although the formulas in Table 3.1 apply also to the wave vector and the four momentum.and can be used in each case also according to the active interpretation, the various situations have their individual features, some of which will now be surveyed.

Consider at first a plane wave the direction of propagation of which makes an angle θ with the direction x_3 of the Lorentz transformation. We write the phase, Equation 3.3.11 of Section 3.2, as

$$\frac{1}{2}\left[(k_0 + k_3)(x_0 - x_3) + (k_0 - k_3)(x_0 + x_3)\right] - k_1 x_1 - k_2 x_2 \tag{3.3.16}$$

This expression is invariant if $(k_0 \pm k_3)$ transforms by the same factor $\exp(\pm\mu)$ as $(x_0 \pm x_3)$.

Thus we have

$$
\begin{aligned}
k_3' &= k_3 \cosh\mu - k_0 \sinh\mu && (3.3.17)\\
k_0' &= -k_3 \sinh\mu + k_0 \cosh\mu && (3.3.18)
\end{aligned}
$$

Since $(k_0, \vec{k})$ is a null-vector, i.e., k has vanishing length, we set

$$
k_3 = k_0 \cos\theta, \quad k_3' = k_0' \cos\theta' \tag{3.3.19}
$$

and we obtain for the aberration and the Doppler effect:

$$
\cos\theta' = \frac{\cos\theta\cosh\mu - \sinh\mu}{\cosh\mu - \cos\theta\sinh\mu} = \frac{\cos\theta - \beta}{1 - \beta\cos\theta} \tag{3.3.20}
$$

and

$$
\frac{k_0'}{k_0} = \frac{\omega_0'}{\omega_0} = \cosh\mu - \cos\theta\sinh\mu \tag{3.3.21}
$$

For $\cos\theta = 1$ we have

$$
\frac{\omega_0'}{\omega_0} = \exp(-\mu) = \sqrt{\frac{1-\beta}{1+\beta}} \tag{3.3.22}
$$

Thus the hyperbolic angle is directly connected with the frequency scaling in the Doppler effect.

Next, we turn to the transformation of the four-momentum of a massive particle. The new feature is that such a particle can be brought to rest. Let us say the particle is at rest in the frame Σ' (rest frame), that moves with the velocity $v_3 = c\tanh^{-1}\mu$ in the frame Σ (lab frame). Thus v_3 can be identified as the particle velocity along x_3.

Solving for the momentum in Σ:

$$
\begin{aligned}
p_3 &= p_3' \cosh\mu + p_0' \sinh\mu && (3.3.23)\\
p_0 &= p_3' \sinh\mu + p_0' \cosh\mu && (3.3.24)
\end{aligned}
$$

with $p_3' = 0, p_0' = mc$, we have

$$
\begin{aligned}
p_3 &= mc\sinh\mu = \frac{mc\beta}{\sqrt{1-\beta^2}} && (3.3.25)\\
p_0 &= mc\cosh\mu = \frac{mc}{\sqrt{1-\beta^2}} = \frac{E}{c}
\end{aligned}
$$

$$
\gamma = \cosh\mu, \qquad \gamma\beta = \sinh\mu \tag{3.3.26}
$$

Thus we have solved the problem posed at the end of Section 3.2.

The point in the preceding argument is that we achieve the transition from a state of rest of a particle to a state of motion, by the kinematic means of inertial transformation. Evidently, the same effect can be achieved by means of acceleration due to a force, and consider this "boost" as an active Lorentz transformation. Let us assume that the particle carries the charge e and is exposed to a constant electric intensity E. We get from Equantion 3.3.25 for small velocities:

$$\frac{dp_3}{dt} = mc\cosh\mu\frac{d\mu}{dt} \simeq mc\frac{d\mu}{dt} \tag{3.3.27}$$

and this agrees with the classical equation of motion if

$$E = \frac{mc}{e}\frac{d\mu}{dt} \tag{3.3.28}$$

Thus the electric intensity is proportional to the hyperbolic angular velocity.

In close analogy, a circular motion can be produced by a magnetic field:

$$B = -\frac{mc}{e}\frac{d\phi}{dt} = -\frac{mc}{e}\omega \tag{3.3.29}$$

This is the well known cyclotron relation.

The foregoing results are noteworthy for a number of reasons. They suggest a close connection between electrodynamics and the Lorentz group and indicate how the group theoretical method provides us with results usually obtained by equations of motion.

All this brings us a step closer to our program of establishing much of physics within a group theoretical framework, starting in particular with the Lorentz group. However, in order to carry out this program we have to generalize our technique to three spatial dimensions. For this we have the choice between two methods.

The first is to represent a four-vector as a 4×1 column matrix and operate on it by 4×4 matrices involving 16 real parameters among which there are ten relations (see Section 2.5).

The second approach is to map four-vectors on Hermitian 2×2 matrices

$$P = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix} \tag{3.3.30}$$

and represent Lorentz transformations as

$$P' = VPV^{\dagger} \tag{3.3.31}$$

where V and $V^{\dagger}$ are Hermitian adjoint unimodular matrices depending .just on the needed six parameters.

We choose the second alternative and we shall show that the mathematical parameters have the desired simple physical interpretations. In particular we shall arrive at generalizations of the de Moivre relation, Equation 3.3.7.

The balance of this chapter is devoted to the mathematical theory of the 2×2 matrices with physical applications to electrodynamics following in Section 4.

3.4 The Pauli Algebra

3.4.1 Introduction

Let us consider the set of all 2×2 matrices with complex elements. The usual definitions of matrix addition and scalar multiplication by complex numbers establish this set as a four-dimensional vector space over the field of complex numbers $\mathcal{V}(4, C)$. With ordinary matrix multiplication, the vector space becomes, what is called an *algebra*, in the technical sense explained at the end of Section 2.3. The nature of matrix multiplication ensures that this algebra, to be denoted $\mathcal{A}_2$, is *associative* and *noncommutative*, properties which are in line with the group-theoretical applications we have in mind.

The name "Pauli algebra" stems, of course, from the fact that $\mathcal{A}_2$ was first introduced into physics by Pauli, to fit the electron spin into the formalism of quantum mechanics. Since that time the application of this technique has spread into most branches of physics.

From the point of view of mathematics, $\mathcal{A}_2$ is merely a special case of the algebra $\mathcal{A}_n$ of $n \times n$ matrices, whereby the latter are interpreted as transformations over a vector space $\mathcal{V}(n^2, C)$. Their reduction to canonical forms is a beautiful part of modern linear algebra.

Whereas the mathematicians do not give special attention to the case $n = 2$, the physicists[13], dealing with four-dimensional space-time, have every reason to do so, and it turns out to be most rewarding to develop procedures and proofs for the special case rather than refer to the general mathematical theorems. The technique for such a program has been developed some years ago[14].

The resulting formalism is closely related to the algebra of complex quaternions, and has been called accordingly a system of hypercomplex numbers. The study of the latter goes back to Hamilton, but the idea has been considerably developed in recent years[15]. The suggestion that the matrices (1) are to be considered symbolically as generalizations of complex numbers which still retain "number-like" properties, is appealing, and we shall make occasional use of it. Yet it seems confining to make this into the central guiding principle. The use of matrices harmonizes better with the usual practice of physics and mathematics[16]

In the forthcoming systematic development of this program we shall evidently cover much ground that is well known, although some of the proofs and concepts of Whitney and Tisza do not seem to be used elsewhere. However, the main distinctive feature of the present approach is that we do not apply the formalism to physical theories assumed to be given, but develop the geometrical, kinematic and dynamic applications in close parallel with the building up of the formalism.

[13] However, see [Ebe65].

[14] [Whi68]. Also unpublished reports by Tisza and Whitney.

[15] See particularly a series of papers by J. D. Edmonds: [Edm73a, Edm75, Edm74b, Edm74a, Edm73b, Edm72]. Also, [Jam74] for the references to the early literature.

[16] For a development of the matrix method see also [Fro75].

Since our discussion is meant to be self-contained and economical, we use references only sparingly. However, at a later stage we shall state whatever is necessary to ease the reading of the literature.

3.4.2 Basic Definitions and Procedures

We consider the set $\mathcal{A}_2$ of all 2×2 complex matrices

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{3.4.1}$$

Although one can generate $\mathcal{A}_2$ from the basis

$$e_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.4.2}$$
$$e_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{3.4.3}$$
$$e_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{3.4.4}$$
$$e_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.4.5}$$

in which case the matrix elements are the expansion coefficients, it is often more convenient to generate it from a basis formed by the Pauli matrices augmented by the unit matrix.

Accordingly $\mathcal{A}_2$ is called the *Pauli algebra*. The basis matrices are

$$\sigma_0 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.4.6}$$
$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.4.7}$$
$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{3.4.8}$$
$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.4.9}$$

The three Pauli matrices satisfy the well known multiplication rules

$$\sigma_j^2 = 1 \quad j = 1, 2, 3 \tag{3.4.10}$$
$$\sigma_j \sigma_k = -\sigma_k \sigma_j = i\sigma_l \quad j\,k\,l = 1\,2\,3 \text{ or an even permutation thereof} \tag{3.4.11}$$

All of the basis matrices are Hermitian, or self-adjoint:

$$\sigma_\mu^\dagger = \sigma_\mu \quad \mu = 0, 1, 2, 3 \tag{3.4.12}$$

(By convention, Roman and Greek indices will run from one to three and from zero to three, respectively.)

We shall represent the matrix A of Equation 3.4.1 as a linear combination of the basis matrices with the coefficient of σ_μ denoted by a_μ. We shall refer to the numbers a_μ as the components of the matrix A. As can be inferred from the multiplication rules, Equation 3.4.11 , matrix components are obtained from matrix elements by means of the relation

$$a_\mu = \frac{1}{2} Tr\left(A\sigma_\mu\right) \tag{3.4.13}$$

where Tr stands for trace. In detail,

$$a_0 = \frac{1}{2}\left(a_{11} + a_{22}\right) \tag{3.4.14}$$

$$a_1 = \frac{1}{2}\left(a_{12} + a_{21}\right) \tag{3.4.15}$$

$$a_2 = \frac{1}{2}\left(a_{12} - a_{21}\right) \tag{3.4.16}$$

$$a_3 = \frac{1}{2}\left(a_{11} - a_{22}\right) \tag{3.4.17}$$

In practical applications we shall often see that a matrix is best represented in one context by its components, but in another by its elements. It is convenient to have full flexibility to choose at will between the two. A set of four components a_μ, denoted by $\{a_\mu\}$, will often be broken into a complex scalar a_0 and a complex "vector"[17] $\{a_1, a_2, a_3\} = \vec{a}$. Similarly, the basis matrices of $\mathcal{A}_2$ will be denoted by $\sigma_0 = 1$ and $\{\sigma_1, \sigma_2, \sigma_3\} = \vec{\sigma}$. With this notation,

$$A = \sum_\mu a_\mu \sigma_\mu = a_0 1 + \vec{a} \cdot \vec{\sigma} \tag{3.4.18}$$

$$= \begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix} \tag{3.4.19}$$

We associate with .each matrix the half trace and the determinant

$$\frac{1}{2} TrA = a_0 \tag{3.4.20}$$

$$|A| = a_0^2 - \vec{a}^2 \tag{3.4.21}$$

[17]The term "vector" here merely signifies a three-component object with which, in a formal way, one can perform the dot and cross products of vector calculus. The term does not refer to the transformation properties to be taken up in Section 3.4.3.

The extent to which these numbers specify the properties of the matrix A, will be apparent from the discussion of their invariance properties in the next two subsections. The positive square root of the determinant is in a way the norm of the matrix. Its nonvanishing: $|A| \neq 0$, is the criterion for A to be invertible.

Such matrices can be normalized to become unimodular:

$$A \to |A|^{-1/2} A \tag{3.4.22}$$

The case of singular matrices

$$|A| = a_0^2 - \vec{a}^2 = 0 \tag{3.4.23}$$

calls for comment. We call matrices for which $|A| = 0$, but $A \neq 0$, null-matrices. Because of their occurrence, $\mathcal{A}_2$ is not a division algebra. This is in contrast, say, with the set of real quaternions which is a division algebra, since the norm vanishes only for the vanishing quaternion.

The fact that null-matrices are important,stems partly from the indefinite Minkowski metric. However, entirely different applications will be considered later.

We list now some practical rules for operations in $\mathcal{A}_2$, presenting them in terms of matrix components rather than the more familiar matrix elements.

To perform matrix multiplications we shall make use of a formula implied by the multiplication rules, Equantion 3.4.11:

$$(\vec{a} \cdot \vec{\sigma}) \left(\vec{b} \cdot \vec{\sigma}\right) = \vec{a} \cdot \vec{b} I + i \left(\vec{a} \times \vec{b}\right) \cdot \vec{\sigma} \tag{3.4.24}$$

where $\vec{a}$ and $\vec{b}$ are complex vectors.

Evidently, for any two matrices A and B

$$[A, B] = AB - BA = 2i \left(\vec{a} \times \vec{b}\right) \cdot \vec{\sigma} \tag{3.4.25}$$

The matrices A and B **commute**, if and only if

$$\vec{a} \times \vec{b} = 0 \tag{3.4.26}$$

that is, if the vector parts $\vec{a}$ and $\vec{b}$ are "parallel" or at least one of them vanishes.

In addition to the internal operations of addition and multiplication, there are external operations on $\mathcal{A}_2$ as a whole, which are analogous to complex conjugation. The latter operation is an *involution*, which means that $(z^*)^* = z$. Of the three involutions any two can be considered independent.

In $\mathcal{A}_2$ we have two independent involutions which can be applied jointly to yield a third:

$$A \quad \to \quad A = a_0 I + \vec{a} \cdot \vec{\sigma} \tag{3.4.27}$$

$$A \quad \to \quad A^\dagger = a_0^* I + \vec{a}^* \cdot \vec{\sigma} \tag{3.4.28}$$

$$A \quad \to \quad \tilde{A} = a_0 I - \vec{a} \cdot \vec{\sigma} \tag{3.4.29}$$

$$A \quad \to \quad \tilde{A}^\dagger = \bar{A} = a_0^* I - \vec{a}^* \cdot \vec{\sigma} \tag{3.4.30}$$

The matrix $A^\dagger$ is the Hermitian adjoint of A. Unfortunately, there is neither an agreed symbol, nor a term for $\tilde{A}$. Whitney called it *Pauli conjugate*, other terms are *quaternionic conjugate* or *hyper-conjugate* $A^\ddagger$ (see Edwards, l.c.). Finally $\bar{A}$ is called *complex reflection*.

It is easy to verify the rules

$$
\begin{aligned}
(AB)^\dagger &= B^\dagger A^\dagger && (3.4.31)\\
\left(\tilde{AB}\right) &= \tilde{B}\tilde{A} && (3.4.32)\\
\left(\bar{AB}\right) &= \bar{B}\bar{A} && (3.4.33)
\end{aligned}
$$

According to Equantion 3.4.33 the operation of complex reflection maintains the product relation in $\mathcal{A}_2$, it is an automorphism. In contrast, the Hermitian and Pauli conjugations are anti-automorphic.

It is noteworthy that the three operations $\tilde{}$, $\dagger$, $\bar{}$, together with the identity operator, form a group (the four-group, "Vierergruppe"). This is a mark of closure: we presumably left out no important operator on the algebra.

In various contexts any of the three conjugations appears as a generalization of ordinary complex conjugation [18].

Here are a few applications of the conjugation rules.

$$A\tilde{A} = \left(a_0^2 - \vec{a}^2\right) 1 = |A| 1 \tag{3.4.34}$$

For invertible matrices

$$A^{-1} = \frac{\tilde{A}}{|A|} \tag{3.4.35}$$

For unimodular matrices we have the useful rule:

$$A^{-1} = \tilde{A} \tag{3.4.36}$$

A Hermitian marrix $A = A^\dagger$ has real components $h_0, \vec{h}$. We define a matrix to be *positive* if it is Hermitian and has a positive trace and determinant:

$$h_0 > 0, \quad |H| = \left(h_0^2 - \vec{h}^2\right) > 0 \tag{3.4.37}$$

If H is positive and unimodular, it can be parametrized as

$$H = \cosh(\mu/2) 1 + \sinh(\mu/2)\hat{h}\cdot\vec{\sigma} = \exp\left\{(\mu/2)\,\hat{h}\cdot\vec{\sigma}\right\} \tag{3.4.38}$$

[18] In the literature one often considers the conjugation of matrix, by taking the complex conjugates of the *elements*: $a_{ik} \rightarrow a_{ik}^*$. This is the case in the well known spinor formalism of van der Waerden. We shall discuss the relation to this formalism in connection with relativistic spinors. However we note already that this convention is asymmetric in the sense that $\sigma_1^* = \sigma_1$, $\sigma_2^* = -\sigma_2$.

The matrix exponential is defined by a power series that reduces to the trigonometric expression. The factor 1/2 appears only for convenience in the next subsection.

In the Pauli algebra, the usual definition $U^\dagger = U^{-1}$ for a *unitary* matrix takes the form

$$u_0^* 1 + \vec{u}^* \cdot \vec{\sigma} = |U|^{-1} (u_0 1 - \vec{u} \cdot \vec{\sigma}) \tag{3.4.39}$$

If U is also unimodular, then

$$u_0^* = u_0 = \text{real} \tag{3.4.40}$$
$$\vec{u}^* = \vec{u} = \text{imaginary} \tag{3.4.41}$$

and

$$\begin{aligned} u_0^2 - \vec{u} \cdot \vec{u} &= u_0^2 + \vec{u} \cdot \vec{u}^* = 1 \\ U = \cos(\phi/2) 1 - i \sin(\phi/2) \hat{u} \cdot \vec{\sigma} &= \exp\left(-i(\phi/2)\hat{u} \cdot \vec{\sigma}\right) \end{aligned} \tag{3.4.42}$$

A unitary unimodular matrix can be represented also in terms of elements

$$U = \begin{pmatrix} \xi_0 & -\xi_1^* \\ \xi_1 & \xi_0^* \end{pmatrix} \tag{3.4.43}$$

with

$$|\xi_0|^2 + |\xi_1|^2 = 1 \tag{3.4.44}$$

where ξ_0, ξ_1, are the so-called Cayley-Klein parameters. We shall see that both this form, and the axis-angle representation, Equation 3.4.42, are useful in the proper context.

We turn now to the category of normal matrices N defined by the condition that they commute with their Hermitian adjoint: $N^\dagger N = N N^\dagger$. Invoking the condition, Equation 3.4.26 , we have

$$\vec{n} \times \vec{n}^* = 0 \tag{3.4.45}$$

implying that n^* is proportional to n, that is all the components of $\vec{n}$ must have the same phase. Normal matrices are thus of the form

$$N = n_0 1 + n \hat{n} \cdot \vec{\sigma} \tag{3.4.46}$$

where n_0 and n are complex constants and $hatn$ is a real unit vector, which we call the axis of N. In particular, any unimodular normal matrix can be expressed as

$$N = \cosh(\kappa/2) 1 + \sinh(\kappa/2) \hat{n} \cdot \vec{\sigma} = \exp\left((\kappa/2)\hat{n} \cdot \vec{\sigma}\right) \tag{3.4.47}$$

where $\kappa = \mu - i\phi, -\infty < \mu < \infty, 0 \leq \phi < 4\pi$, and $\hat{n}$ is a real unit vector. If $\hat{n}$ points in the “3” direction, we have

$$N_0 = \exp\left[\left(\frac{\kappa}{2}\right) \sigma_3\right] = \begin{pmatrix} \exp(\frac{\kappa}{2}) & 0 \\ 0 & \exp(-\frac{\kappa}{2}) \end{pmatrix} \tag{3.4.48}$$

Thus the matrix exponentials, Equations 3.4.38, 3.4.42 and 3.4.48, are generalizations of a diagonal matrix and the latter is distinguished by the requirement that the axis points in the z direction.

Clearly the normal matrix, Equation 3.4.48, is a commuting product of a positive matrix like Equation 3.4.38 with $\hat{h} = \hat{n}$ and a unitary matrix like Equation 3.4.42, with $\hat{u} = \hat{n}$:

$$N = HU = UH \tag{3.4.49}$$

The expressions in Equation 3.4.49 are called the *polar forms* of N, the name being chosen to suggest that the representation of N by H and U is analogous to the representation of a complex number z by a positive number r and a phase factor:

$$z = r\exp(-i\phi/2) \tag{3.4.50}$$

We shall show that, more generally, *any invertible matrix* has two unique polar forms

$$A = HU = UH' \tag{3.4.51}$$

but only the polar forms of normal matrices display the following equivalent special features:

1. H and U commute
2. $\hat{h} = \hat{u} = \hat{n}$
3. $H' = H$

We see from the polar decomposition theorem that our emphasis on positive and unitary matrices is justified, since all matrices of $\mathcal{A}_2$ can be produced from such factors. We proceed now to prove the theorem expressed in Equation 3.4.51 by means of an explicit construction.

First we form the matrix $AA^\dagger$, which is positive by the criteria 3.4.36:

$$a_0 a_0^* + \vec{a}\cdot\vec{a}^* > 0 \tag{3.4.52}$$

$$|A||A^\dagger| > 0 \tag{3.4.53}$$

Let $AA^\dagger$ be expressed in terms of an axis $\hat{h}$ and the hyperbolic angle μ:

$$\begin{aligned} AA^\dagger &= b\left(\cosh\mu 1 + \sinh\mu\hat{h}\cdot\hat{\sigma}\right) \\ &= b\exp(\mu\hat{h}\cdot\hat{\sigma}) \end{aligned} \tag{3.4.54}$$

where b is a positive constant. We claim that the Hermitian component of A is the positive square root of 3.4.54

$$H = \left(AA^\dagger\right)^{1/2} = b^{1/2}\exp(\frac{\mu}{2}\hat{h}\cdot\hat{\sigma}) \tag{3.4.55}$$

with

$$U = H^{-1}A, \quad A = HU \tag{3.4.56}$$

That U is indeed unitary is easily verified:

$$U^\dagger = A^\dagger H^{-1}, \quad U^{-1} = A^{-1}H \tag{3.4.57}$$

and these expressions are equal by Equation 3.4.55.

From Equation 3.4.56 we get

$$A = U(U^{-1}HU)$$

and

$$A = UH' \quad \text{with} \quad H' = U^{-1}HU \tag{3.4.58}$$

It remains to be shown that the polar forms 3.4.56 are unique. Suppose indeed, that for a particular A we have two factorizations

$$A = HU = H_1U_1 \tag{3.4.59}$$

then

$$AA^\dagger = H^2 = H_1^2 \tag{3.4.60}$$

But, since $AA^\dagger$ has a unique positive square root, $H_1 = H$, and

$$U = H_1^{-1}A = H^{-1}A = U \quad \textit{q.e.d.} \tag{3.4.61}$$

Polar forms are well known to exist for any $n \times n$ matrix, although proofs of uniqueness are generally formulated for abstract transformations rather than for matrices, and require that the transformations be invertable [19].

3.4.3 The restricted Lorentz group

Having completed the classification of the matrices of $\mathcal{A}_2$, we are ready to interpret them as operators and establish a connection with the Lorentz group. The straightforward procedure would be to introduce a 2-dimensional complex vector space $\mathcal{V}(\in, \mathcal{C})$. By using the familiar bra-ket formalism we write

$$A|\xi\rangle = |\xi'\rangle \tag{3.4.62}$$

$$A^\dagger\langle\xi| = \langle\xi'| \tag{3.4.63}$$

The two-component complex vectors are commonly called spinors. We shall study their properties in detail in Section 5. The reason for this delay is that the physical interpretation of spinors is a

[19] See [Hal58], page 170, and [Gel61], page 111. The theorem is stated only partially by Halmos as $A = UH'$, and by Gel'fand in the weak form $A = HU = U'H'$ without stating $U = U'$. See also [HK61], page 343. For an extension of the theorem to singular matrices see [Whi71].

subtle problem with many ramifications. One is well advised to consider at first situations in which the object to be operated upon can be represented by a 2×2 matrix.

The obvious choice is to consider Hermitian matrices, the components of which are interpreted as relativistic four-vectors. The connection between four-vectors and matrices is so close that it is often convenient to use the same symbol for both:

$$A = a_0 1 + \vec{a} \cdot \vec{\sigma} \tag{3.4.64}$$

$$A = \{a_0, \vec{a}\} \tag{3.4.65}$$

We have

$$a_0^2 - \vec{a}^2 = |A| = \frac{1}{2} Tr(A\bar{A}) \tag{3.4.66}$$

or more generally

$$a_0 b_0 - \vec{a} \cdot \vec{b} = \frac{1}{2} Tr(A\bar{B}) \tag{3.4.67}$$

A Lorentz transformation is defined as a linear transformation

$$\{a_0, \vec{a}\} = \mathcal{L}\{a_0', \vec{a}'\} \tag{3.4.68}$$

that leaves the expression 3.4.67 and hence also 3.4.66 invariant. We require moreover that the sign of the "time component" a_0 be invariant (orthochronic Lorentz transformation $L^{\uparrow}$), and that the determinant of the $4x4$ matrix $\mathcal{L}$ be positive (proper Lorentz transformation L_+). If both conditions are satisfied, we speak of the restricted Lorentz group $L_+^{\uparrow}$. This is the only one to be of current interest for us, and until further notice "Lorentz group" is to be interpreted in this restricted sense.

Note that A can be interpreted as any of the four-vectors discussed in Section 3.2: $R = \{r, \vec{r}\}$,

$$K = \{k_0, \vec{k}\}, \quad P = \{p_0, \vec{p}\} \tag{3.4.69}$$

Although these vectors and their matrix equivalents have identical transformation properties, they differ in the possible range of their determinants. A negative $|P|$ can arise only for an unphysical imaginary rest mass. By contrast, a positive R corresponds to a time-like displacement pointing toward the future, an R with a negative $|R|$ to a space-like displacement and $|R| = 0$ is associated with the light cone. For the wave vector we have by definition $|K| = 0$.

To describe a Lorentz transformation in the Pauli algebra we try the "ansatz"

$$A' = VAW \tag{3.4.70}$$

with $|V| = |W| = 1$ in order to preserve $|A|$. Reality of the vector, i.e., hermiticity of the matrix A is preserved if the additional condition $W = V^{\dagger}$ is satisfied. Thus the transformation

$$A' = VAV^{\dagger} \tag{3.4.71}$$

leaves expression 3.4.66 invariant. It is easy to show that 3.4.67 is invariant as well.

The complex reflection $\bar{A}$ transforms as

$$\bar{A}' = \bar{V}\bar{A}\tilde{V} \tag{3.4.72}$$

and the product of two four-vectors:

$$\begin{aligned}(A\bar{B})' &= VAV^{\dagger}\bar{V}\bar{B}\tilde{V} \\ &= V(A\bar{B})V^{-1}\end{aligned} \tag{3.4.73}$$

This is a so-called similarity transformation. By taking the trace of Equation 3.4.73 we confirm that the inner product 3.4.67 is invariant under 3.4.72. We have to remember that a cyclic permutation does not affect the trace of a product of matrices [20]. Thus Equation 3.4.72 indeed induces a Lorentz transformation in the four-vector space of A.

It is well known that the converse statement is also true: to every transformation of the restricted Lorentz group $L_+^{\uparrow}$ there are associated two matrices differing-only by sign (their parameters ϕ differ by 2π) in such a fashion as to constitute a two-to-one homomorphism between the group of unimodular matrices $\mathcal{SL}(2,C)$ and the group $L_+^{\uparrow}$. It is said also that $\mathcal{SL}(2,C)$ provides a two-valued representation of $L_+^{\uparrow}$. We shall prove this statement by demonstrating explicitly the connection between the matrices V and the induced, or associated group operations.

We note first that A and $\bar{A}$ correspond in the tensor language to the contravariant and the covariant representation of a vector. We illustrate the use of the formalism by giving an explicit form for the inverse of 3.4.72

$$A = V^{-1}A'V^{\dagger -1} \equiv \tilde{V}A'\bar{V} \tag{3.4.74}$$

We invoke the polar decomposition theorem Equation 3.4.49 of Section 3.4.2 and note that it is sufficient to establish this connection for unitary and positive matrices respectively.

Consider at first

$$A' = UAU^{\dagger} \equiv UAU^{-1} \tag{3.4.75}$$

with

$$\begin{aligned}U(\hat{u}, \frac{\phi}{2}) &\equiv \exp\left(-\frac{i\phi}{2}\hat{u}\cdot\vec{\sigma}\right) \\ u_1^2 + u_2^2 + u_3^2 = 1, &\qquad 0 \le \phi < 4\pi\end{aligned} \tag{3.4.76}$$

The set of all unitary unimodular matrices described by Equation 3.4.76 form a group that is commonly called $\mathcal{SU}(2)$.

[20] Since for hermitian matrices $\tilde{A} = \bar{A}$, we use now $\bar{A}$ as in Equation 3.4.34 (Section 3.4.2), since in Section 4 $\tilde{A}$ will prove to be of greater generality, and being automorphic, it is easier to handle.

Let us decompose $\vec{a}$:

$$\vec{a} = \vec{a}_{\parallel} + \vec{a}_{\perp} \tag{3.4.77}$$

$$\vec{a}_{\parallel} = (\vec{a}\cdot\hat{u})\hat{u}, \qquad \vec{a}_{\perp} = \vec{a} - \vec{a}_{\parallel} = \hat{u}\times(\vec{a}\times\hat{u}) \tag{3.4.78}$$

It is easy to see that Equation 3.4.75 leaves a_0 and $a_{\parallel}$ invariant and induces a rotation around $\hat{u}$ by an angle ϕ: $R\{\hat{u},\phi\}$.

Conversely, to every rotation $R\{\hat{u},\phi\}$ there correspond two matrices:

$$U(\hat{u},\frac{\phi}{2}) \quad \text{and} \quad U(\hat{u},\frac{\phi+2\pi}{2}) = -U(\hat{u},\frac{\phi}{2}) \tag{3.4.79}$$

We have $1 \to 2$ homomorphism between $\mathcal{SO}(3)$ and $\mathcal{SU}(2)$, the latter is said to be a two-valued representation of the former. By establishing this correspondence we have solved the problem of parametrization formulated on page 13. The nine parameters of the orthogonal 3×3 matrices are reduced to the three independent ones of $U(\hat{u},\frac{\phi}{2})$. Moreover we have the simple result

$$U^n = \exp\left(-\frac{in\phi}{2}\hat{u}\cdot\vec{\sigma}\right) \tag{3.4.80}$$

which reduces to the de Moivre theorem if $\hat{n}\cdot\vec{\sigma} = \sigma_3$.

Some comment is in order concerning the two-valuedness of the $\mathcal{SU}(2)$ representation. This comes about because of the use of half angles in the algebraic formalism which is deeply rooted in the geometrical structure of the rotation group. (See the Rodrigues-Hamilton theorem in Section 2.2.)

Whereas the two-valuedness of the $\mathcal{SU}(2)$ representation does not affect the transformation of the A vector based on the bilateral expression 3.4.75, the situation will be seen to be different in the spinorial theory based on Equation 3.4.62, since under certain conditions the sign of the spinor $|\xi\rangle$ is physically meaningful [21].

The above discussion of the rotation group is incomplete even within the classical theory. The rotation $R\{\hat{u},\phi\}$ leaves vectors along $\hat{u}$ unaffected. A more appropriate object to be rotated is the Cartesian triad, to be discussed in Section 5.

We consider now the case of a positive matrix $V = H$

$$A' = HAH \tag{3.4.81}$$

with

$$H = \exp(\frac{\mu}{2}\hat{h}\cdot\sigma) \tag{3.4.82}$$

$$h_1^2 + h_2^2 + h_3^2 = 1, \qquad -\infty < \mu < \infty \tag{3.4.83}$$

[21] Historically, $\mathcal{SU}(2)$ was introduced into physics in order to account for the electron spin within quantum mechanics. This is a case where the two-valuedness of the formalism is significant, although not too well understood. For this reason there is a tendency in the literature to work as long as possible with the single-valued representations involving 3×3 matrices. On closer inspection this turns out to be a false economy.

We decompose $\vec{a}$ as

$$\vec{a} = a\hat{h} + \vec{a}_{\perp} \tag{3.4.84}$$

and using the fact that $(\vec{a} \cdot \vec{\sigma})$ and $(\vec{b} \cdot \vec{\sigma})$ commute for $\vec{a} \| \vec{b}$ and anticommute for $\vec{a} \perp \vec{b}$, we obtain

$$A' = \exp(\frac{\mu}{2}\hat{h} \cdot \sigma) \left(a_0 1 + a\hat{h} \cdot \sigma + \vec{a}_{\perp} \cdot \sigma \right) \exp(\frac{\mu}{2}\hat{h} \cdot \sigma) \tag{3.4.85}$$

$$= \exp(\mu\hat{h} \cdot \sigma) \left(a_0 1 + \vec{a}\hat{h} \cdot \sigma \right) + \vec{a}_{\perp} \cdot \sigma \tag{3.4.86}$$

Hence

$$a_0' = \cosh \mu a_0 + \sinh \mu a \tag{3.4.87}$$

$$a' = \sinh \mu a_0 + \cosh \mu a_0 \tag{3.4.88}$$

$$\vec{a}_{\perp}' = \vec{a}_{\perp} \tag{3.4.89}$$

This is to be compared with Table 3.1, but remember that we have shifted from the passive to the active interpretation, from *alias* to *alibi*.

Positive matrices with a common axis form a group (Wigner's "little group"), but in general the product of Hermitian matrices with different axes are not Hermitian. There arises a unitary factor, which is the mathematical basis for the famous Thomas precession.

Let us consider now a normal matrix

$$V = N = H(\hat{k}, \frac{\mu}{2})U(\hat{k}, \frac{\phi}{2}) = \exp\left(\frac{\mu - i\phi}{2} \hat{n} \cdot \sigma \right) \tag{3.4.90}$$

where we have the commuting product of a rotation and a Lorentz transformation with the same axis $\hat{n}$. Such a constellation is called a Lorentz 4-screw[22].

An arbitrary sequence of pure Lorentz transformations and pure rotations is associated with a pair of matrices V and $-V$, which in the general case is of the form

$$H(\hat{h}, \frac{\mu}{2})U(\hat{u}, \frac{\phi}{2}) = U(\hat{u}, \frac{\phi}{2})H'(\hat{h}', \frac{\mu}{2}) \tag{3.4.91}$$

According to Equation 3.4.58 of Section 3.4.2, H and H' are connected by a similarity transformation, which does not affect the angle μ, but only the axis of the transformation. (See the next section.)

This matrix depends on the 6 parameters, $\hat{h}, \mu, \hat{u}, \phi$, and thus we have solved the general problem of parametrization mentioned above.

For a normal matrix $\hat{h} = \hat{u} = \hat{n}$ and the number of parameters is reduced to 4.

[22] See [Syn65], page 89.

Our formalism enables us to give a closed form for two arbitrary normal matrices and the corresponding 4-screws.

$$[N, N'] = 2i \sinh\frac{\kappa}{2} \sinh\frac{\kappa}{2} \,(\hat{n} \times \hat{n}') \cdot \vec{\sigma} \tag{3.4.92}$$

where $\kappa = \mu - i\phi$, $\kappa' = \mu' - i\phi'$.

In the literature the commutation relations are usually given in terms of infinitesimal operators which are defined as follows:

$$\begin{aligned}
U(\hat{u}_k, \frac{d\phi}{2}) &= 1 - \frac{i}{2} d\phi\sigma_k = 1 + d\phi I_k && (3.4.93)\\
I_k &= -\frac{i}{2}\sigma_k && (3.4.94)\\
H(\hat{h}_k, \frac{d\mu}{2}) &= 1 + \frac{d\mu}{2}\sigma_k = 1 + d\mu L_k && (3.4.95)\\
L_k &= \frac{1}{2}\sigma_k && (3.4.96)
\end{aligned}$$

The commutation relations are

$$\begin{aligned}
[I_1, I_2] &= I_3 && (3.4.97)\\
[L_1, L_2] &= -I_3 && (3.4.98)\\
[L_1, I_2] &= L_3 && (3.4.99)
\end{aligned}$$

and cyclic permutations.

It is a well known result of the Lie-Cartan theory of continuous group that these infinitesimal-generators determine the entire group. Since we have represented these generators in $\mathcal{SL}(2, C)$, we have completed the demonstration that the entire group $L_+^\uparrow$ is accounted for in our formalism.

3.4.4 Similarity classes and canonical forms of active transformations

It is evident that a Lorentz transformation induced by a matrix $H(\hat{h}, \frac{\mu}{2})$ assumes a particularly simple form if the z-axis of the coordinate system is placed in the direction of $\hat{h}$. The diagonal matrix $H(\hat{z}, \frac{\mu}{2})$ is said to be the canonical form of the transformation. This statement is a special case of the problem of canonical forms of linear transformations, an important chapter in linear algebra.

Let us consider a linear mapping in a vector space. A particular choice of basis leads to a matrix representation of the mapping, and representations associated with different frames are connected by *similarity transformations*. Let A_1, be an arbitrary and S an invertible matrix. A similarity transformation is effected on A, by

$$A_2 = SA_1S^{-1} \tag{3.4.100}$$

Matrices related by similarity transformation are called *similar*, and matrices similar to each other constitute a *similarity class*.

In usual practice the mapping-refers to a vector space as in Equation 3.4.62 of Section 3.4.3:

$$A_1|\xi\rangle_1 = |\xi'\rangle_1 \tag{3.4.101}$$

The subscript refers to the basis "1." A change of basis $\Sigma_1 \to \Sigma_2$ is expressed as

$$|\xi\rangle_2 = S|\xi\rangle_1, \quad |\xi'\rangle_2 = S|\xi'\rangle_1 \tag{3.4.102}$$

Inserting into Equation 3.4.101 we obtain

$$A_1 S^{-1}|\xi\rangle_2 = S^{-1}|\xi\rangle_2 \tag{3.4.103}$$

and hence

$$A_2|\xi\rangle_2 = |\xi\rangle_2 \tag{3.4.104}$$

where A_2 is indeed given by Equation 3.4.100.

The procedure we have followed thus far to represent Lorentz transformations in $\mathcal{A}_2$ does not quite follow this standard pattern.

We have been considering mappings of the space of fourvectors which in turn were represented as 2×2 complex matrices. Thus both operators and operands are matrices of $\mathcal{A}_2$. In spite of this difference in interpretation, the matrix representations in different frames are still related according to Equation 3.4.100.

This can be shown as follows. Consider a unimodular matrix A, that induces a Lorentz transformation in P-space, whereby the matrices refer to the basis Σ_1:

$$P_1' = A_1 P_1 A_1^\dagger \tag{3.4.105}$$

We interpret Equation 3.4.105 in the active sense as a *linear mapping* of P-space on itself that corresponds physically to some *dynamic process* that alters P in a linear way.

We shall see in Section 4 that the Lorentz force acting on a charged particle during the time dt can be indeed considered as an active Lorentz transformation. (See also page 26.)

The process has a physical meaning independent of the frame of the observer, but the matrix representations of P, P' and of A depend on the frame. The four-momenta in the two frames are connected by a Lorentz transformation interpreted in the passive sense:

$$P_2 = S P_1 S^\dagger \tag{3.4.106}$$

$$P_2 = S P_1' S^\dagger \tag{3.4.107}$$

with $|S| = 1$. Solving for P, P' and inserting into Equation 3.4.105, we obtain

$$S^{-1} P_2' \tilde{S}^\dagger = A_1 S^{-1} P_2 \tilde{S}^\dagger A_1^\dagger S \tag{3.4.108}$$

or

$$P_2' = A_2 P_2 A_1^\dagger \tag{3.4.109}$$

where A_2 and A_1 are again connected by the similarity transformation 3.4.100.

We may apply the polar decomposition theorem to the matrix S. In the special case that S is unitary, we speak of a unitary similarity transformation corresponding to the rotation of the coordinate system discussed at the onset of this section. However, the general case will lead us to less obvious physical applications.

The above considerations provide sufficient motivation to examine the similarity classes of $\mathcal{A}_2$. We shall see that all of them have physical applications, although the interpretation of singular mappings will be discussed only later.

The similarity classes can be characterized in several convenient ways. For example, one may use two independent similarity invariants shared by all the matrices $A = a_0 l + \vec{a} \cdot \vec{\sigma}$ in the class. We shall find it convenient to choose

1. the determinant $|A|$, and
2. the quantity $\vec{a}^2$

The trace is also a similarity invariant, but it is not independent: $a_0^2 = |A| + \vec{a}^2$.

Alternatively, one can characterize the whole class by one representative member of it, some matrix A_0 called the *canonical form* for the class (See Table 3.2).

We proceed at first to characterize the similarity classes in terms of the invariants 1 and 2. We recall that a matrix A is invertible if $|A| \neq 0$ and singular if $|A| = 0$. Without significant loss of generality, we can normalize the invertible matrices of $\mathcal{A}_2$ to be unimodular, so that we need discuss only classes of singular and of unimodular matrices. As a second invariant to characterize a class, we choose $\vec{a} \cdot \vec{a}$, and we say that a matrix A is axial if $\vec{a} \cdot \vec{a} \neq 0$. In this case, there exists a unit vector $\hat{a}$ (possibly complex) such that $\vec{a} = a \cdot \hat{a}$ where a is a complex constant. The unit vector $\hat{a}$ is called the axis of A. Conversely, the matrix A is non-axial if $\vec{a} \cdot \vec{a} = 0$, the vector $\vec{a}$ is called *isotropic* or a *null-vector*, it cannot be expressed in terms of an axis.

The concept of axis as here defined is the generalization of the real axis introduced in connection with normal matrices on page 33. The usefulness of this concept is apparent from the following theorem:

Theorem 1. For any two unit vectors $\hat{v}_1$, and $\hat{v}_2$, real or complex, there exists a matrix S such that

$$\hat{v}_2 \cdot \vec{\sigma} = S \hat{v}_1 \cdot \vec{\sigma} S^1 \tag{3.4.110}$$

Proof. We construct one such matrix S from the following considerations. If $\hat{v}_1$, and $\hat{v}_2$ are real, then let S be the unitary matrix that rotates every vector by an angle π about an axis which bisects the angle between $\hat{v}_1$, and $\hat{v}_2$:

$$S = -i\hat{s} \cdot \vec{\sigma} \tag{3.4.111}$$

where

$$\hat{s} = \frac{\hat{v}_1 + \hat{v}_2}{\sqrt{2\hat{v}_1 \cdot \hat{v}_2 + 2}} \tag{3.4.112}$$

Even if $\hat{v}_1$, and $\hat{v}_2$ are not real, it is easily verified that S as given formally by Equations 3.4.111 and 3.4.112, does indeed send $\hat{v}_1$ to $\hat{v}_2$. Naturally S is not unique; for instance, any matrix of the form

$$S = \exp\left\{\left(\frac{\mu_2}{2} - i\frac{\phi_2}{2}\right)\vec{v}_2 \cdot \vec{\sigma}\right\}(-i\hat{s}\cdot\vec{\sigma})\exp\left\{\left(\frac{\mu_1}{2} - i\frac{\phi_1}{2}\right)\vec{v}_1 \cdot \vec{\sigma}\right\} \tag{3.4.113}$$

will send $\hat{v}_1$ to $\hat{v}_2$.

This construction fails only if

$$\hat{v}_1 \cdot \hat{v}_2 + 1 = 0 \tag{3.4.114}$$

that is for the transformation $\hat{v}_1 \rightarrow -\hat{v}_2$. In this trivial case we choose

$$S = -i\hat{s}\cdot\vec{\sigma}, \quad \text{where} \quad \hat{s} \perp \vec{v}_1 \tag{3.4.115}$$

□

Since in the Pauli algebra diagonal matrices are characterized by the fact that their axis is $\hat{x}_3$, we have proved the following theorem:

Theorem 2. All axial matrices are diagonizable, but normal matrices and only normal matrices are diagonizable by a unitary similarity transformation.

The diagonal forms are easily ascertained both for singular and the unimodular cases. (See Table 3.2.) Because of their simplicity they are called also *canonical forms*. Note that they can be multiplied by any complex number in order to get all of the axial matrices of $\mathcal{A}_2$.

The situation is now entirely clear: the canonical forms show the nature of the mapping; a unitary similarity transformation merely changes the geometrical orientation of the axis. The angle of circular and hyperbolic rotation specified by a_0 is invariant. A general transformation complexifies the axis. This situation comes about if in the polar form of the matrix $A = HU$, the factors have distinct real axes, and hence do not commute.

There remains to deal with the case of nonaxial matrices. Consider $A = \vec{a}\cdot\vec{\sigma}$ with $\vec{a}^2 = 0$. Let us decompose the isotropic vector $\vec{a}$ into real and imaginary parts:

$$\vec{a} = \vec{\alpha} + i\vec{\beta} \tag{3.4.116}$$

Hence $\vec{\alpha}^2 - \vec{\beta}^2 = 0$ and $\alpha \cdot \beta = 0$. Since the real and the imaginary parts of a are perpendicular, we can rotate these directions by a unitary similarity transformation into the x- and y-directions respectively. The transformed matrix is

$$\frac{\alpha}{2}(\sigma_1 + i\sigma_2) = \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} \tag{3.4.117}$$

with a positive. A further similarity transformation with

$$S = \begin{pmatrix} \alpha^{-1/2} & 0 \\ 0 & \alpha^{1/2} \end{pmatrix} \tag{3.4.118}$$

transforms Equation 3.4.117 into the canonical form given in Table 3.2.

As we have seen in Section 3.4.3 all unimodular matrices induce Lorentz transformations in Minkowski, or four-momentum space. According to the results summarized in Table 3.2, the mappings induced by axial matrices can be brought by similarity transformations into so-called Lorentz four-screws consisting of a circular and hyperbolic rotation around the same axis, or in other words: a rotation around an axis, and a boost along the same axis.

What about the Lorentz transformation induced by a nonaxial matrix? The nature of these transformations is very different from the common case, and constitutes an unusual limiting situation. It is justified to call it an exceptional Lorentz transformation. The special status of these transformations was recognized by Wigner in his fundamental paper on the representations of the Lorentz group[23].

The present approach is much more elementary than Wigner's, both from the point of view of mathematical technique, and also the purpose in mind. Wigner uses the standard algebraic technique of elementary divisors to establish the canonical Jordan form of matrices. We use, instead a specialized technique adapted to the very simple situation in the Pauli algebra. More important, Wigner was concerned with the problem of representations of the inhomogeneous Lorentz group, whereas we consider the much simpler problem of the group structure itself, mainly in view of application to the electromagnetic theory.

The intuitive meaning of the exceptional transformations is best recognized from the polar form of the generating matrix. This can be carried out by direct application of the method discussed at the end of the last section. It is more instructive, however, to express the solution in terms of (circular and hyperbolic) trigonometry.

We ask for the conditions the polar factors have to satisfy in order that the relation

$$1 + \hat{a} \cdot \vec{\sigma} = H(\hat{h}, \frac{\mu}{2}) U(\hat{u}, \frac{\phi}{2}) \tag{3.4.119}$$

should hold with $\mu \neq 0$, $\phi \neq 0$. Since all matrices are unimodular, it is sufficient to consider the equality of the traces:

$$\frac{1}{2} Tr A = \cosh(\frac{\mu}{2}) \cos(\frac{\phi}{2}) - i \sinh(\frac{\mu}{2}) \sin(\frac{\phi}{2})\, \hat{h} \cdot \hat{u} = 1 \tag{3.4.120}$$

This condition is satisfied if and only if

$$\hat{h} \cdot \hat{u} = 0 \tag{3.4.121}$$

[23][Wig39]; See also [Syn65], particularly for the connections with the electromagnetic field treated in our Section 4. [Pae69], p. 114, This author speaks of "α-transformations."

and

$$\cosh(\frac{\mu}{2})\cos(\frac{\phi}{2}) = 1 \tag{3.4.122}$$

The axes of circular and hyperbolic rotation are thus perpendicular, to each other and the angles of these rotations are related in a unique fashion: half of the circular angle is the so-called Gudermannian function of half of the hyperbolic angle

$$\frac{\phi}{2} = gd\left(\frac{\mu}{2}\right) \tag{3.4.123}$$

However, if μ and ϕ are infinitesimal, we get

$$\left(1 + \frac{\mu^2}{2} + \dots\right)\left(1 + \frac{\phi^2}{2} + \dots\right) = 1, \text{i.e.,} \tag{3.4.124}$$

$$\mu^2 - \phi^2 = 0 \tag{3.4.125}$$

We note finally that products of exceptional matrices need not be exceptional, hence exceptional Lorentz transformations do not form a group.

In spite of their special character, the exceptional matrices have interesting physical applications, both in connection with the electromagnetic field as discussed in Section 4, and also for the construction of representations of the inhomogeous Lorentz group [Pae69, Wig39].

We conclude by noting that the canonical forms of Table 3.2 lend themselves to express the powers A_0^k in simple form.

For the axial singular matrix we have

$$A_0^2 = A \tag{3.4.126}$$

These *projection matrices* are called *idempotent*. The nonaxial singular matrices are *nilpotent*:

$$A_0^2 = 0 \tag{3.4.127}$$

The exceptional matrices (unimodular nonaxial) are raised to any power k (even non-real) by the formula

$$A^k = 1^k\left(1 + k\vec{a}\cdot\vec{\sigma}\right) \tag{3.4.128}$$

$$= 1^k \exp\left(k\vec{a}\cdot\vec{\sigma}\right) \tag{3.4.129}$$

For integer k, the factor 1^k becomes unity. The axial unimodular case is handled by formulas that are generalizations of the well known de Moivre formulas:

$$A^k = 1^k \exp\left(k\frac{\kappa}{2} + kl2\pi i\right) \tag{3.4.130}$$

where l is an integer. For integer k, Equation 3.4.130 reduces to

$$A^k = \exp\left(k(\frac{\kappa}{2})\vec{a}\cdot\vec{\sigma}\right) \tag{3.4.131}$$

In connection with these formulae, we note that for *positive* A ($\phi = 0$ and a real) , there is a unique positive mth root of A:

$$
\begin{aligned}
A &= \exp\left\{(\frac{\mu}{2})\hat{a}\cdot\vec{\sigma}\right\} && (3.4.132)\\
A^{1/m} &= \exp\left\{(\frac{\mu}{2m})\hat{a}\cdot\vec{\sigma}\right\} && (3.4.133)
\end{aligned}
$$

The foregoing results are summarized in Table 3.2.

	Unimodular $\lvert A\rvert = 1$	Singular $\lvert A\rvert = 0$
Axial $\vec{a} = a\hat{a}$	$A = \exp(\frac{\kappa}{2}\vec{a}\cdot\vec{\sigma})$ $A_0 = \begin{pmatrix} \exp(\frac{\kappa}{2}) & 0 \\ 0 & \exp(\frac{-\kappa}{2}) \end{pmatrix}$ $\kappa = \mu - i\phi$ $-\infty < \mu < \infty$ $0 \le \phi < 4\pi$	$A = \frac{1}{2}(1 \pm \vec{a}\cdot\vec{\sigma})$ $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
Non-axial $\vec{a}^2 = 0$ $\vec{a} = \frac{1}{2}(\hat{e}_1 \pm \hat{e}_2)$	$A = 1 + \vec{a}\cdot\vec{\sigma} = \exp(\vec{a}\cdot\vec{\sigma})$ $A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$	$A = \vec{a}\cdot\vec{\sigma}$ $A_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

Table 3.2: Canonical Forms for the Simlarity classes of $\mathcal{A}_2$.