

Notes on Physics from Symmetry

Author: Juvid Aryaman

Last compiled: October 17, 2021

This document contains my personal notes on Jakob Schwichtenberg's Physics from Symmetry ([Schwichtenberg, 2015](#)), with a sprinkling of notes from my undergraduate physics course in quantum field theory (and, to a lesser extent, general relativity).

1. Special relativity

1.1. Definitions and postulates

In special relativity, **inertial frames of reference** are coordinate systems moving with constant velocity relative to each other. Special relativity has two basic postulates:

1. **The principal of relativity:** The laws of physics are the same in all inertial frames of reference.
2. **The invariance of the speed of light:** The velocity of light has the same value c in all inertial frames of reference.

Theorem 1.1 (Invariant of special relativity). *Consider two events A and B in an inertial observer O 's frame of reference. Let the time interval measured by O between the two events be (Δt) , and the three spatial intervals be (Δx) , (Δy) , (Δz) . Then, the quantity*

$$(\Delta s)^2 := (\Delta ct)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \quad (1.1)$$

is invariant between all frames of reference. I.e.

$$(\Delta s')^2 = (\Delta s)^2 \quad (1.2)$$

for any inertial frame of reference O' .

Theorem 1.1 follows directly from the invariance of the speed of light (consider a pair of mirrors, for two observers with relative velocity).

Definition 1.1 (Proper time). *Proper time, τ , is the time measured by an observer in the special frame of reference where the object in question is at rest. In this frame of reference,*

$$(\Delta s)^2 = (c\Delta\tau)^2. \quad (1.3)$$

In the infinitesimal limit

$$(ds)^2 = (cd\tau)^2. \quad (1.4)$$

Physically, Defn. 1.1 means that all observers agree on the time interval between events for an observer who travels with the object in question. However, different observers **do not** in general agree on the time interval between events generally: $(\Delta t) \neq (\Delta t')$ – this is called **time dilation**.

1.2. c is an upper speed limit

All observers agree on the value of $(ds)^2 = (cd\tau)^2$. Furthermore, we commonly assume that there exists a minimal proper time of $\tau = 0$ for two events if $\Delta s^2 = 0$. We can therefore write that when $\tau = 0$

$$c^2 = \frac{(dx)^2 + (dy)^2 + (dz)^2}{(dt)^2} \quad (1.5)$$

between two events with an infinitesimal distance. We can equate the right-hand side with a squared velocity, and hence

$$\tau = 0 \implies c^2 = v^2 \quad (1.6)$$

so

$$(ds)^2 \geq 0 \implies c^2 \geq v^2 \quad (1.7)$$

for **any** pair of events (which are causally connected, although how this follows is not immediately clear to me right now).

1.3. Einstein notation

Moving from left to right, the first index denotes a row, the second index denotes a column, and so on for higher-order tensors. Hence

$$(\mathbf{A})_{ij} := A^i_j \quad (1.8)$$

where the right hand side is Einstein notation, and the left hand side is usual matrix element notation.

Throughout this document, where it is instructive to do so, we use a bold symbol (e.g. \mathbf{A}) to stress where we are using conventional matrix notation. We denote the value of the i th row and j th column of the matrix as $(\mathbf{A})_{ij}$. However, it is cumbersome and unconventional in the physics community to always denote matrices with a bold symbol, so we often drop the bold symbol and simply denote a matrix \mathbf{A} as simply A . We will, however, **avoid** writing things like $A_{\mu\nu}$ to denote the (μ, ν) th element of an arbitrary matrix because such notation has a very particular meaning in Einstein notation. If you see something like $A_{\mu\nu}$ (i.e. a normal typeface with indices) in this document, then you should assume that we mean what we say¹ – in this case $A_{\mu\nu} = A^\rho_\nu g_{\rho\mu}$ which is **not** the same as $(\mathbf{A})_{\mu\nu}$ in general (see below for explanation).

In Einstein notation, we can use the mnemonic “Upper indices go up to down [column vector, contravariant]; lower indices go left to right [row vector, covariant]”. We commonly use the conventions:

- Greek letters are used for spacetime components, where indices take values 0, 1, 2, or 3.
- Latin letters are used for spatial components only, where indices take values 1, 2, or 3.

The **Einstein summation convention** states that where an index appears twice in a single term, and not otherwise defined, it implies summation of that term over all values of the index.

Matrix-vector multiplication The product of a matrix A_{ij} with a column vector v_j is written as

$$u^i = A^i_j v^j. \quad (1.9)$$

Matrix multiplication For a product of two matrices

$$C^i_k = A^i_j A^j_k \iff (\mathbf{C})_{ik} = (\mathbf{A}\mathbf{B})_{ik}. \quad (1.10)$$

The upstairs/downstairs-ness of this equation is irrelevant: what matters is the left/right **ordering** of the indices, as well as which indices are repeated. For example

$$A^\mu_\delta \eta_{\delta\nu} = (\mathbf{A}\boldsymbol{\eta})_{\mu\nu}. \quad (1.11)$$

¹Unless, of course, I’ve made a mistake! In which case, my dear reader, you could leave a [git issue](#) and I’ll do my best to make a fix.

Transpose The transpose of a matrix is written as

$$(\mathbf{B}^T)_{ij} := (\mathbf{B})_{ji} = B^j_i \quad (1.12)$$

So, for example

$$A^\delta_\mu \eta_{\delta\nu} = (A^T)^\mu_\delta \eta_{\delta\nu} = (\mathbf{A}^T \boldsymbol{\eta})_{\mu\nu}. \quad (1.13)$$

Raising/lowering with metric tensor In order to be able to freely raise/lower indices (without repeatedly writing the metric tensor), we impose an ordering upon indices of tensor fields – which we represent typographically with spacing between tensor indices. A metric g_{ij} (or g^{ij}) has the effect of lowering (or raising) a repeated index. For example,

$$g_{iq} T^{abcd}_{efgh}{}^{ijkl}_{mnop} = T^{abcd}_{efghq}{}^{ijkl}_{mnop}. \quad (1.14)$$

1.4. Minkowski spacetime

Definition 1.2 (Four-vector (contravariant)). A position four-vector is defined as

$$x^\mu = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}. \quad (1.15)$$

Definition 1.3 (Minkowski metric). The Minkowski metric is defined as

$$\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1). \quad (1.16)$$

η is used to compute distances and lengths in Minkowski space.

We define $\eta^{\mu\nu}$ through the relation

$$\eta^{\mu\nu} \eta_{\nu\sigma} = \delta^\mu_\sigma. \quad (1.17)$$

Definition 1.4 (One-form (covariant vector)). We define a one-form as

$$x_\mu = \eta_{\mu\nu} x^\nu. \quad (1.18)$$

Thus,

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu. \quad (1.19)$$

Definition 1.5 (Scalar product). A scalar product between four-vectors x and y is defined as

$$x \cdot y := x^\mu y^\nu \eta_{\mu\nu} = x_\mu y_\nu \eta^{\mu\nu} = x^\mu y_\mu = x_\nu y^\nu \quad (1.20)$$

due to the symmetry of the metric: $\eta_{\mu\nu} = \eta_{\nu\mu}$.

1.5. Lorentz transformations

From the invariant of SR (Theorem 1.1), we have

$$ds'^2 = dx'_\mu dx'_\nu \eta^{\mu\nu} = dx_\mu dx_\nu \eta^{\mu\nu} \quad (1.21)$$

for all reference frames. We denote Λ as a (1,1) tensor field, which transforms a four-vector from one reference frame to another:

$$dx'^\mu = \Lambda^\mu_\nu dx^\nu \quad (1.22)$$

which leaves the ds^2 invariant, i.e. $ds'^2 = ds^2$. It follows that

$$\eta_{\mu\nu} = \Lambda^\sigma{}_\mu \Lambda^\delta{}_\nu \eta_{\sigma\delta}. \quad (1.23)$$

We can write this in ordinary matrix notation using Eq.(1.12)

$$\begin{aligned} \eta_{\mu\nu} &= \Lambda^\sigma{}_\mu \eta_{\sigma\delta} \Lambda^\delta{}_\nu \\ \eta_{\mu\nu} &= (\Lambda^T)^\mu{}_\sigma \eta_{\sigma\delta} \Lambda^\delta{}_\nu \\ (\boldsymbol{\eta})_{\mu\nu} &= (\boldsymbol{\Lambda}^T \boldsymbol{\eta} \boldsymbol{\Lambda})_{\mu\nu} \\ \boldsymbol{\eta} &= \boldsymbol{\Lambda}^T \boldsymbol{\eta} \boldsymbol{\Lambda}. \end{aligned}$$

The physical meaning of Eq.(1.23) is that Lorentz transformations leave the scalar product of Minkowski spacetime invariant: i.e. changes between frames of reference that respect the two postulates of special relativity (Section 1.1). Conservation of the scalar product is analogous to rigid rotation (O) in Euclidean space ($a \cdot b = a' \cdot b' = a^T O^T O b \implies O^T I O = I$), which preserves orientation ($\det(O) = 1$).

Note that $\Lambda^\mu{}_\nu \neq \Lambda_\nu{}^\mu$. Beginning with Eq.(1.23),

$$\Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma \eta_{\mu\nu} = \eta_{\rho\sigma}$$

we can raise one index, and lower one index, of $\Lambda^\nu{}_\sigma$

$$\begin{aligned} \Lambda^\mu{}_\rho \eta_{\mu\nu} \Lambda^\nu{}_\sigma \eta^{\sigma\lambda} &= \eta_{\rho\sigma} \eta_{\nu\mu} \eta^{\sigma\lambda} \\ \Lambda^\mu{}_\rho \Lambda_\mu{}^\lambda \eta_{\nu\sigma} &= \eta_{\nu\sigma} \delta_\rho{}^\lambda \\ \Lambda^\mu{}_\rho \Lambda_\mu{}^\lambda &= \delta_\rho{}^\lambda \end{aligned} \quad (1.24)$$

so we see that $\Lambda_\nu{}^\mu$ is the inverse of $\Lambda^\mu{}_\nu$.

2. Lie group theory

2.1. Invariance, symmetry, and covariance

We call a quantity **invariant** if it does not change under particular transformations. E.g. if we transform $A, B, C, \dots \rightarrow A', B', C', \dots$ and we have

$$F(A', B', C', \dots) = F(A, B, C, \dots) \quad (2.1)$$

then we say F is invariant under this transformation. **Symmetry** is defined as invariance under a transformation (or class of transformations). An equation is covariant if it takes the same form when objects in it are transformed. *All physical laws must be covariant under Lorentz transformations.*

Group theory describes the properties of particular sets of transformations: the invariances under such groups allows us to mathematically describe symmetry. For example, the set of rotations about the origin of a square by $n\pi/2$ form a **discrete group**, and leave the set of points which constitute the square invariant under the transformation. The set of rotations about the origin of a circle form a **continuous group**. We can use group theory to work with *all* kinds of symmetries: symmetries which operate on vectors, equations, ...

2.2. Groups

Definition 2.1 (Group axioms). A group (G, \circ) is a set G , together with a binary operation \circ defined on G , that satisfies the following axioms

- **Closure:** For all $g_1, g_2 \in G$, $g_1 \circ g_2 \in G$

- *Identity element:* There exists an identity element $e \in G$ such that for all $g \in G$, $g \circ e = g = e \circ g$
- *Inverse element:* For each $g \in G$, there exists an inverse element $g^{-1} \in G$ such that $g \circ g^{-1} = e = g^{-1} \circ g$.
- *Associativity:* For all $g_1, g_2, g_3 \in G$, $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$

The set of all transformations that leave a given object invariant is called a **symmetry group**. For Minkowski spacetime, the object that is left invariant is the Minkowski metric, and the corresponding symmetry group is called the **Poincaré group**. Notice that the transformations which constitute a group are defined entirely independently from the object on which the transformations act.

2.2.1. Rotations in two dimensions and $SO(2)$

Consider the 2D rotation matrix

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (2.2)$$

and the two reflection matrices

$$P_x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad P_y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.3)$$

These matrices satisfy the group axioms. We can uncover this group from a symmetry perspective. The above transformations leave the length of a vector unchanged, i.e.

$$a \cdot a = a' \cdot a'. \quad (2.4)$$

Letting the transformation be represented by $a' = Oa$, it follows that all members of the group must satisfy

$$O^T O = I. \quad (2.5)$$

This condition defines the group $O(2)$, which is the group of all **orthogonal** 2×2 matrices. It follows that $\det(O) = \pm 1$ – i.e. the transformations are area-preserving. The subgroup with $\det(O) = 1$ is called $SO(2)$, which corresponds to rigid rotations preserving the orientation of the system – “S” denoting **special**.

2.2.2. Rotations with unit complex numbers and $U(1)$

A unit complex number is a complex number z which satisfies $|z|^2 = z^* z = 1$. The group $U(1)$ is the set of unit complex numbers, together with ordinary complex number multiplication. The U stands for ‘**unitary**’, which generally stands for the condition

$$U^\dagger U = 1, \quad (2.6)$$

where $U^\dagger = (U^T)^*$ is the **Hermitian conjugate** of U . For scalars, the Hermitian conjugate is equivalent to the complex conjugate. Note that a unit complex number can also be denoted as

$$R_\theta = e^{i\theta} = \cos(\theta) + i \sin(\theta) \quad (2.7)$$

which makes the interpretation of $U(1)$ as rotations on the unit complex numbers evident.

We can connect this description of rotations ($U(1)$) to the previous ($SO(2)$) by defining

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (2.8)$$

For an arbitrary unit complex number $z = a + ib$, let

$$f(z) = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \quad (2.9)$$

Since $z = R_\theta = \cos(\theta) + i \sin(\theta)$, we can plug in the real and imaginary components of z into Eq.(2.9) to arrive at Eq.(2.2). We then have $z' = R_\theta z$, to perform rotations. There therefore exists an **isomorphism** between $SO(2)$ and $U(1)$:

Definition 2.2 (Group isomorphism). Given two groups $(G, *)$, (H, \odot) , a group isomorphism is a bijective function $f : G \rightarrow H$ such that

$$f(u * v) = f(u) \odot f(v) \quad \forall u, v \in G. \quad (2.10)$$

which is written as

$$(G, *) \cong (H, \odot). \quad (2.11)$$

$f(z)$ in Eq.(2.9) is therefore a group isomorphism between $U(1)$ and $SO(2)$,

$$SO(2) \cong U(1). \quad (2.12)$$

This realization has an analogue in three dimensions, which will reveal something fundamental about nature.

2.2.3. Rotations in three dimensions and $SO(3)$

As with $SO(2)$, the defining conditions of $SO(3)$ are

$$R^T R = I \quad (2.13)$$

$$\det(R) = 1. \quad (2.14)$$

Matrices satisfying these two conditions represent rigid rotations of 3-dimensional vectors. These matrices can be described by the following three “basis rotations”

$$\begin{aligned} R_x &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} & R_y &= \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \\ R_z &= \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (2.15)$$

So, to rotate a vector v around the z -axis by θ , we would compute $R_z(\theta)v$. The set of (orientation-preserving) rotation matrices acting on 3-dimensional vectors is called $SO(3)$.

2.2.4. Quaternions and $SU(2)$

To get a second description of rotations in three dimensions, we must generalize complex numbers in higher dimensions. Astonishingly, it turns out that there are no 3-dimensional complex numbers. Instead, we can find 4-dimensional complex numbers called quaternions, which will turn out to be able to describe rotations in 3-dimensions. The fact that quaternions are 4-dimensional will reveal something deep about the universe. We could have anticipated this result, because we will be using unit quaternions, which have 3 degrees of freedom.

×	1	i	j	k
1	1	i	j	k
i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

Figure 1. Quaternion multiplication table, read as row \times column = value. E.g. $\mathbf{j}\mathbf{i} = -\mathbf{k}$. In general, the basic quaternions anti-commute.

To construct quaternions, we introduce three complex units satisfying the relations

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 \quad (2.16)$$

$$\mathbf{ijk} = -1 \quad (2.17)$$

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}. \quad (2.18)$$

All other relations can be computed from the above. For example, the relation $\mathbf{ij} = \mathbf{k}$ can be derived by multiplying both sides of Eq.(2.18) by \mathbf{k} . Notice that it follows that the **basic quaternions anticommute** with each other, see Fig. 1.

The set of unit quaternions satisfy

$$q^\dagger q = 1 \quad (2.19)$$

$$\implies a^2 + b^2 + c^2 + d^2 = 1. \quad (2.20)$$

As the unit complex numbers formed a group under complex number multiplication, the unit quaternions form a group under quaternion multiplication. There are several possible ways of representing the basic quaternions with 2D matrices, but one way is as follows:

$$\begin{aligned} \mathbf{1} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \mathbf{i} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ \mathbf{j} &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, & \mathbf{k} &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \end{aligned} \quad (2.21)$$

With these matrices, a generic quaternion $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ can be written in a matrix representation as

$$f(q) = \begin{pmatrix} a + di & b + ci \\ -b + ci & a - di \end{pmatrix}. \quad (2.22)$$

We also observe that $\det(f(q)) = 1$, and so we conclude that the unit quaternions are given by the set of matrices with the above form and unit determinant. The unit quaternions, written as 2×2 matrices U therefore fulfil the conditions

$$U^\dagger U = 1 \quad \text{and} \quad \det(U) = 1. \quad (2.23)$$

This defines the symmetry group $SU(2)$.

The map between $SU(2)$ and $SO(3)$ is not as simple as the one we saw between $U(1)$ and $SO(2)$. The mapping of a complex number onto a 2-dimensional vector is easy because a complex number has two degrees of freedom: $v = x + iy$. But the mapping of a quaternion onto 3-dimensional vector is not so straightforward because a quaternion has four degrees of freedom. We will make the mapping of a 3-dimensional vector $(x, y, z)^T$ onto a quaternion v as

$$v \equiv x\mathbf{i} + y\mathbf{j} + z\mathbf{k}. \quad (2.24)$$

Using Eq.(2.21), we see that $\det(v) = x^2 + y^2 + z^2$. In order to perform transformations which preserve the length of the vector (x, y, z) , we must use transformations which preserve determinants. Therefore, the restriction to *unit* quaternions means that we must restrict to matrices with *unit* determinants². Naively, a first guess would be that simply multiplying a vector v by a unit quaternion u induces a rotation on v , but this is not the case because the product of u and v may not belong to $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$. It turns out that the following transformation can describe rotations in 3-dimensions

$$v' = qvq^{-1}. \quad (2.25)$$

Let t be a quaternion defining a rotation through ϕ , where

$$t = \cos\left(\frac{\phi}{2}\right) + \sin\left(\frac{\phi}{2}\right)u \quad (2.26)$$

$$u = u_x\mathbf{i} + u_y\mathbf{j} + u_z\mathbf{k} \quad (2.27)$$

$$u^\dagger u = 1 \implies t^\dagger t = 1. \quad (2.28)$$

As an example, suppose we wish to rotate the vector $\vec{v} = (1, 0, 0)^T$ around the z -axis by ϕ . Then using Eq.(2.21)

$$\vec{v} = (1, 0, 0)^T \rightarrow v = 1\mathbf{i} + 0\mathbf{j} + 0\mathbf{k} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (2.29)$$

From Eq.(2.26), defining $\theta = \phi/2$

$$R_z(\theta) = \cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{k} = \begin{pmatrix} \cos(\theta) + i\sin(\theta) & 0 \\ 0 & \cos(\theta) - i\sin(\theta) \end{pmatrix}. \quad (2.30)$$

From Eq.(2.25), the rotated vector v' is

$$v' = R_z(\theta)vR_z(\theta)^{-1} = \begin{pmatrix} 0 & \cos(\phi) + i\sin(\phi) \\ -\cos(\phi) + i\sin(\phi) & 0 \end{pmatrix}. \quad (2.31)$$

Using the general quaternion matrix representation Eq.(2.22), we can equate

$$v'_x = \cos(\phi), \quad v'_y = \sin(\phi), \quad v'_z = 0 \quad (2.32)$$

as expected.

Inspection of Eq.(2.26) reveals that the mapping of unit quaternions onto 3-dimensional rotations is not one-to-one. For example, a rotation by $\phi = \pi$ is equivalent to a rotation by $\phi = 2\pi + \pi = 3\pi$. But,

$$t_{\phi=\pi} = \sin\left(\frac{\pi}{2}\right)u = u \quad (2.33)$$

$$t_{\phi=3\pi} = \sin\left(\frac{3\pi}{2}\right)u = -u. \quad (2.34)$$

Hence, we call $SU(2)$ a **double-cover** of $SO(3)$, because every element of $SO(3)$ has two corresponding elements in $SU(2)$ [**TODO: I think?**]. It is therefore always possible to go unambiguously from $SU(2)$ to $SO(3)$, but not vice versa.

We will see later that groups which cover other groups are fundamental for quantum spin. Note that the above had one quaternion parameter too many, which may be interpreted as a hint towards relativity: a more natural identification may have been $v = t\mathbf{1} + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Rotations in 4-dimensions require 6 degrees of freedom³. However, there is no 7-dimensional generalisation of complex numbers. But two unit quaternions do have 6 degrees of freedom: we will see later how there is a close connection between two copies of $SU(2)$ and rotations in four dimensions.

²Since $\det(BA) = \det(B)\det(A)$

³Using ordinary 4×4 matrices, with the constraints $O^T O = 1$ and $\det(O) = 1$ reduce the 16 components of an arbitrary 4×4 matrix to 6 independent components

3. Lie algebras

3.1. Intuition behind generators

Consider an element of a continuous group which is arbitrarily close to the identity

$$g(\epsilon) = I + \epsilon X \quad (3.1)$$

where ϵ is small, and I is the identity (think about 1D rotations for concreteness). We let X remain abstract for the moment. We wish to apply this transformation N times to achieve a finite transformation by a total amount θ , $h(\theta)$. As a result, we may write:

$$h(\theta) = (I + \frac{\theta}{N} X)^N. \quad (3.2)$$

We then take the limit of $N \rightarrow \infty$, which is just the exponential function

$$h(\theta) = \lim_{N \rightarrow \infty} (I + \frac{\theta}{N} X)^N = e^{\theta X}. \quad (3.3)$$

In this sense, the object X generates the finite transformation h , which is why we call X a **generator**.

We can then differentiate $h(\theta)$ to obtain the generator:

$$X = \left. \frac{dh(\theta)}{d\theta} \right|_{\theta=0}. \quad (3.4)$$

If we consider a continuous group of transformations that are given by matrices, we can also make a Taylor expansion of an element of the group about the identity

$$h(\theta) = \sum_n \frac{1}{n!} \left. \frac{d^n h}{d\theta^n} \right|_{\theta=0} \theta^n \quad (3.5)$$

which shows how generators generate transformations.

3.2. An intuitive definition for matrix Lie groups

For matrix Lie groups, the corresponding Lie algebra can be defined as the collection of objects that give an element of the group when exponentiated.

Definition 3.1 (Lie algebra for matrix Lie groups, non-rigorous). For a matrix Lie group G (given by $n \times n$ matrices), the Lie algebra \mathfrak{g} of G is given by those $n \times n$ matrices X such that

$$e^{tX} \in G \quad (3.6)$$

for $t \in \mathbb{R}$, together with an operation called the Lie bracket, that defines how to combine elements of the Lie algebra

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}. \quad (3.7)$$

A Lie algebra is **closed** under the Lie bracket.

Note that, in general $X \circ Y \notin \mathfrak{g}$.

Theorem 3.1. Consider a matrix Lie group G and corresponding Lie algebra \mathfrak{g} . If $X, Y \in \mathfrak{g}$ and $g, h \in G$, then

$$g \circ h = e^X \circ e^Y = \underbrace{e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+\dots}}_{\in G} \quad (3.8)$$

where the right hand side is the Baker-Campbell-Hausdorff formula⁴. For **matrix Lie groups**, the Lie bracket $[X, Y]$ is **equivalent** to the **commutator** of X and Y :

$$[X, Y] = XY - YX. \quad (3.9)$$

Eq.(3.8) connects combinations of elements of the group G to combinations of elements of the Lie algebra \mathfrak{g} .

3.3. Generators and Lie algebra of $SO(3)$

Using the norm-preserving condition of $SO(3)$, Eq. (2.13), along with Eq.(3.6) $O = e^{\theta J}$, yields

$$O^T O = e^{\theta J^T} e^{\theta J} = 1 \implies J^T + J = 0, \quad (3.10)$$

The parity-preserving condition in Eq.(2.14) also yields⁵

$$\det(e^{\theta J}) = e^{\theta \text{tr}(J)} = 1 \implies \text{tr}(J) = 0. \quad (3.11)$$

A set of three **basis**⁶ generators J_1, J_2, J_3 fulfilling both of these conditions can be written conveniently as

$$(J_i)_{jk} = -\epsilon_{ijk}, \quad (3.12)$$

where ϵ_{ijk} is the Levi-Civita symbol

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) = \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\} \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } k = i \\ -1 & \text{if } (i, j, k) = \{(1, 3, 2), (2, 1, 3), (3, 2, 1)\} \end{cases}. \quad (3.13)$$

The corresponding Lie bracket is

$$[J_i, J_j] = \epsilon_{ijk} J_k. \quad (3.14)$$

In physics, it is conventional to define the generators of $SO(3)$ with an extra i , so that we get Hermitian generators⁷, to arrive at

$$O = e^{i\phi J} \quad (3.15)$$

$$\implies (J_i)_{jk} = -i\epsilon_{ijk} \quad (3.16)$$

$$[J_i, J_j] = i\epsilon_{ijk} J_k. \quad (3.17)$$

We call the Lie bracket relation of the basis generators **the** Lie algebra of a given group, because everything that is important about a Lie algebra is encoded in the Lie bracket relation of the basis generators.

An alternative way to derive the basis generators is to begin with the finite transformation matrices and apply Eq.(3.4). However, the above method is more general, whose steps were:

1. Begin with the definition of the group;
2. Use those constraints to derive the basis generators;
3. Then derive the an explicit form for finite transformations.

⁴This holds more generally, not only for matrix Lie groups. See also Defn. 3.2

⁵Because $\det(e^A) = e^{\text{tr}(A)}$

⁶i.e. all other generators are linear combinations of the basis generators

⁷i.e. $J^\dagger = J$

3.4. Formal definition of a Lie algebra

Definition 3.2 (Lie algebra). A Lie algebra is a vector space \mathfrak{g} over some field⁸ F together with a binary operation $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ called the Lie bracket satisfying the following axioms:

1. Bilinearity:

$$[aX + bY, Z] = a[X, Z] + b[Y, Z] \quad (3.18)$$

$$[Z, aX + bY] = a[Z, X] + b[Z, Y] \quad (3.19)$$

for all scalars $a, b \in F$ and all elements $X, Y, Z \in \mathfrak{g}$.

2. Anticommutativity:

$$[X, Y] = -[Y, X] \quad (3.20)$$

for all $X \in \mathfrak{g}$.

3. The Jacobi identity

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0 \quad (3.21)$$

for all elements $X, Y \in \mathfrak{g}$.

Importantly, this definition makes no reference to any Lie group: the definition of a Lie algebra stands on its own.

3.5. Generators and Lie algebra of $SU(2)$

Using Equations (2.23) for the definition of $SU(2)$, and Eq.(3.15) for the definition of a generator, we arrive at

$$e^{-iJ^\dagger} e^{iJ_i} = e^{-iJ_i^\dagger + iJ_i + \frac{1}{2}[J_i^\dagger, J_i] + \dots} = 1 \quad (3.22)$$

where the right-hand side uses the Baker-Campbell-Hausdorff theorem. This implies that

$$J = J^\dagger \quad (3.23)$$

$$\text{tr}(J) = 0 \quad (3.24)$$

where the second equation uses $\det(U) = 1$. A basis for Hermitian traceless 2×2 matrices is given by the 3 **Pauli** matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.25)$$

We conventionally define

$$J_i \equiv \sigma_i/2 \quad (3.26)$$

to arrive at the Lie bracket (or, equivalently, commutator relationship)

$$[J_i, J_k] = i\epsilon_{ijk} J_k \quad (3.27)$$

which is **identical** to Eq.(3.17) for $SO(3)$! Hence we say that $SU(2)$ and $SO(3)$ have the same Lie algebra, since they share the same Lie bracket.

⁸Informally, a field is a set, with operations of addition and multiplication defined on that set, along with inverses of those operations

3.6. Lie groups and covering groups

An $SU(2)$ transformation doesn't necessarily have to be represented by 2×2 matrices. An abstract definition of a Lie group will enable us to see the connection between different descriptions of the same transformation.

A **Lie group** is a group that is also a differentiable manifold. A Lie group must have the group operations \circ induce a **differentiable** map of the manifold onto itself. Concretely, this means that every group element, e.g. A , induces a differentiable map that takes any element of the group B to another element of the group $C = AB$.

Definition 3.3 (Simply connected Lie group). A Lie group is said to be simply connected if every closed curve on the manifold can be shrunk smoothly to a point.

Theorem 3.2 (Existence of a covering group). There exists a unique simply connected Lie group corresponding to each Lie algebra. We call this Lie group the **covering group**. All other groups having the same Lie algebra are said to be “covered” by the simply connected Lie group.⁹

Definition 3.4 (Unit n -sphere). A unit n -sphere is denoted as the set of points satisfying

$$S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}. \quad (3.28)$$

S^n is therefore a sphere embedded in $n + 1$ dimensions, but with a unit radius constraint, implying an n -dimensional object. For groups we have discussed so far, we may identify:

1. S^1 , with group operation of multiplication, is the covering group of 1-dimensional rotations. $S^1 \cong U(1) \cong SO(2)$.
2. S^3 , with group operation of multiplication, is the covering group of 3-dimensional rotations. $S^3 \cong SU(2) \xrightarrow{\text{two-to-one}} SO(3)$.

Geometrically, we can think of $SO(3)$ as a 3-sphere but with antipodal points identified: this is because each point of $SO(3)$ is associated with two points of $SU(2)$ which is equivalent to S^3 .

3.7. Representation theory

Representation theory deals with *representing* abstract algebraic structures (mainly groups, associative algebras, and Lie algebras) with linear transformations of vector spaces – i.e. matrices operating on columnar vectors. Considering just groups for now, a representation abides the following definition

Definition 3.5 (Representation for a group). A representation is a map R (or **homomorphism**) between any group element g of a group G onto a vector space V such that:

- $R(e) = I$, i.e. the identity element of the group is also the matrix identity
- $R(g^{-1}) = (R(g))^{-1}$, i.e. inverse elements are mapped to inverse transforms
- $R(g) \circ R(h) = R(gh)$.

For example, the usual rotation matrices in Eq.(2.15) are a representation of the group $SO(3)$ on the vector space \mathbf{R}^3 . Representation theory allows us to examine the effect of the group on other vector spaces – of potentially different dimensionality than the definition of the group itself.

Theorem 3.3 (Similarity transformation). If R is a representation of group G with elements g , then for any invertible matrix S ,

$$R'(g) = S^{-1}R(g)S \quad (3.29)$$

is also a representation.

⁹This is a result from differential geometry.

Proof. Follows immediately from the definition of a representation. For example,

$$R'(g_1)R'(g_2) = S^{-1}R(g_1)SS^{-1}R(g_2)S = S^{-1}R(g_1)R(g_2)S = S^{-1}R(g_1g_2)S = R'(g_1g_2). \quad (3.30)$$

□

This freedom to perform similarity transformations corresponds to the freedom to choose a basis for the vector space the group acts on.

Definition 3.6 (Invariant subspace). Given a representation R of a group G on vector space V , we call $V' \subseteq V$ an invariant subspace if for all $v \in V'$, we have $R(g)v \in V'$.

Definition 3.7 (Subrepresentation). A subrepresentation of a representation (R, V) of a group G is a representation (R', V') such that $V' \subseteq V$ and

$$R'(g)v = R(g)v \quad (3.31)$$

for all $v \in V'$.

Definition 3.8 (Irreducible representation). An irreducible representation is a representation (R, V) of a group G that has no invariant subspace besides the zero space $\{0\}$ and V itself.

Theorem 3.4 (Irreducible representations and block diagonal form). An irreducible representation is a representation which cannot be rewritten, using a similarity transformation, into block diagonal form. In contrast, reducible representations can be rewritten in block diagonal form through similarity transformations.

There are many representations for any particular group. How do we know which ones to choose to describe nature?

Definition 3.9 (General linear group of a vector subspace). $GL(V)$ is the set of $n \times n$ invertible matrices, together with ordinary matrix multiplication on a vector space V . It is the group of automorphisms of V , i.e. the set of bijective linear transformations $V \rightarrow V$.

Lemma 3.1 (Schur's Lemma). If we have an irreducible representation $R : \mathfrak{g} \rightarrow GL(V)$, then any linear operator $T : V \rightarrow V$ that commutes with all operators $R(X)$ for $X \in \mathfrak{g}$, must be a scalar multiple of the identity operator.

Definition 3.10 (Casimir element). A Casimir element C is an object with the property

$$[C, X] = 0 \quad (3.32)$$

for every $X \in \mathfrak{g}$ where \mathfrak{g} is a Lie algebra. Note that if C commutes with each basis of \mathfrak{g} then it follows that C obeys Eq.(3.32).

Consequently, by Schur's Lemma, the Casimir element for any given representation of a Lie algebra must be a scalar multiple of the identity. This scalar provides us with a number to label representations of Lie algebras. **Physical mass and spin turn out to be examples of these constants of proportionality.**

3.8. $SU(2)$

On our way to describing the Poincaré group, we will first discuss irreducible representations of $SU(2)$, which will then allow us to start talking about special relativity again.

3.8.1. Finite dimensional irreducible representations

Definition 3.11 (Quadratic Casimir element). *Quadratic Casimir elements are the simplest Casimir elements and are constructed as quadratic polynomials of generators of a Lie algebra.*

For the Lie algebra $\mathfrak{su}(2)$, there exists a **unique** Casimir element

$$J^2 = J_1^2 + J_2^2 + J_3^2 \quad (3.33)$$

where J_i is the basis for the Lie algebra Eq.(3.26). Different dimensional representations will give different representations of the Casimir operator: in 2-D $J_{2d}^2 = 3/4 \cdot I$ whereas $J_{3d}^2 = 2 \cdot I$. Letting $J^2 v = bv$, for some vector v , then we will use b to label the representation.

Definition 3.12 (Cartan element). *Cartan elements of a representation of a Lie algebra are those basis generators which are simultaneously diagonalizable.¹⁰*

Only one basis generator of $\mathfrak{su}(2)$ can be diagonalized at a time, therefore the Cartan element of $\mathfrak{su}(2)$ is unique. It is customary to use J_3 for this purpose. We will show below that the eigenvalues of J_3 (m) label the elements of the vector space that the Lie algebra acts upon.

Definition 3.13 (Bra-ket notation). *If ϕ is a vector in a complex vector space, then we use $|\phi\rangle$ to denote this vector (pronounced “ket-phi”). Any convenient label can be used inside the ket $|\rangle$, which denotes that the object is simply a vector in a vector space. A bra $\langle\phi|$ is an element of the dual vector space, i.e. a bra is a linear functional which is a linear map from the vector space to the complex numbers.*

We will therefore label every basis vector, for every representation of a Lie algebra, in bra-ket notation as follows

$$\begin{aligned} J^2 |b, m\rangle &= b |b, m\rangle \\ J_3 |b, m\rangle &= m |b, m\rangle \end{aligned} \quad (3.34)$$

where m labels the representation and b labels the vector within the representation. We can obtain basis vectors for the representation by enumerating the eigenvalues of J^2 and J_3 .

Definition 3.14 (Ladder operators for $\mathfrak{su}(2)$). *Define the following complex linear combination of the basis generators of $SU(2)$*

$$J_+ = J_1 + iJ_2 \quad (3.35)$$

$$J_- = J_1 - iJ_2. \quad (3.36)$$

where J_+ is a **raising operator** and J_- is a **lowering operator**.

In $\mathfrak{su}(2)$ we only consider real linear combinations of generators, so we thus expand our consideration to the **complexification** of $\mathfrak{su}(2)$, called $\mathfrak{sl}(2, \mathbb{C})$ where $SL(2, \mathbb{R})$ is the special linear group.

Applying Eq.(3.27) gives rise to two new commutation relations

$$[J_3, J_{\pm}] = \pm J_{\pm} \quad (3.37)$$

$$[J_+, J_-] = 2J_3. \quad (3.38)$$

As a result, we find that if $|b, m\rangle$ is an eigenvector of J_3 with eigenvalue m , then $J_{\pm} |b, m\rangle$ is also an eigenvector of J_3 with eigenvalue $m \pm 1$, i.e.

$$J_3 J_{\pm} |b, m\rangle = (m \pm 1) J_{\pm} |b, m\rangle \quad (3.39)$$

$$\implies J_{\pm} |b, m\rangle \propto |b, m \pm 1\rangle. \quad (3.40)$$

¹⁰A set of matrices are simultaneously diagonalizable if there exists a matrix P such that $P^{-1}AP$ is a diagonal matrix for every A in the set.

There can only be a finite number of eigenvectors of J_3 because we are dealing with a finite-dimensional representation, because the vector space is finite dimensional and therefore there can only be a finite number of linearly independent vectors. Therefore, there must exist an eigenvector with a maximum eigenvalue – let's call it j . It must have the property

$$J_+ |b, j\rangle = 0. \quad (3.41)$$

Using

$$J_- J_+ = J^2 - J_3^2 - J_3 \quad (3.42)$$

then it follows from the action of $J_- J_+ |b, j\rangle$ that

$$b = j(j+1). \quad (3.43)$$

If we let k be the minimum eigenvalue of J_3 , then an analogous argument allows us to conclude that $k = -j$ and therefore

$$-j \leq m \leq j. \quad (3.44)$$

Hence we have derived a relationship between the whole representation and the number of states in the vector space.

Now, if we begin at $|b, j\rangle$ and by applying the lowering operator J_- we reduce the eigenvalue of J_3 by 1. We perform this operation some finite number of times, and hence $j - k$ must be an integer. But $k = -j$, so

$$2j = \text{integer} \implies j = \frac{\text{integer}}{2}. \quad (3.45)$$

Simply plugging in $j = 0, 1/2, 1, 3/2, \dots$ allows us to enumerate all representations of $\mathfrak{su}(2)$.

Lastly, by assuming that our basis vectors $|b, m\rangle$ are normalized

$$|b, m\rangle^\dagger |b, m\rangle = 1 \quad (3.46)$$

we can derive the constant of proportionality in Eq.(3.40). By observing that $(J_+ |b, m\rangle)^\dagger J_+ |b, m\rangle = |C|^2$ where C is the constant of proportionality, and by using an analogous argument for J_- , we arrive at

$$J_+ |j(j+1), m\rangle = \sqrt{j(j+1) - m^2 - m} |j(j+1), m+1\rangle \quad (3.47)$$

$$J_- |j(j+1), m\rangle = \sqrt{j(j+1) - m^2 + m} |j(j+1), m-1\rangle. \quad (3.48)$$

Finally, note that the representation label $j(j+1)$ takes up a lot of space and is somewhat redundant, so we tend to label states simply as $|j, m\rangle$ rather than $|j(j+1), m\rangle$. The tools above will allow us to derive explicit representations of $SU(2)$ in different dimensions.

3.8.2. Recipe for explicit $\mathfrak{sl}(2, \mathbb{C})$ representations of $SU(2)$

First, we choose a particular value of j – the maximum value of the eigenvalue of the Casimir element of the group. In 2-dimensions, this is $j = 1/2$. We then take all the corresponding eigenvalues of the Casimir element, in this case $m = 1/2, -1/2$. We then construct a diagonal matrix for representation of the Cartan element:

$$J_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.49)$$

We therefore have the basis vectors of the Lie algebra as

$$|1/2, 1/2\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1/2, -1/2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.50)$$

We can find the explicit matrix form of the other two $SU(2)$ generators by rewriting Eq.(3.36) in terms of J_1 and J_2 . We can apply J_1 and J_2 on the two basis vectors above, and concatenate the corresponding vectors into the matrix representations of J_1 and J_2 . We then recover (3.26). This can be repeated for any dimension you like.

3.9. The Lorentz group $O(1, 3)$

We now have the tools to derive a finite-dimensional representation of the Lorentz group, of which Lorentz transforms are elements.

The Lorentz group is the set of all transformations which preserve the inner product in Minkowski space

$$x^\mu x_\mu = x^\mu \eta_{\mu\nu} x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \quad (3.51)$$

where $\eta_{\mu\nu}$ is defined in Eq.(1.16). The reason we call the group $O(1, 3)$ is due to the signature of the Minkowski metric: the group $O(4)$ preserves $(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$.

Recall that Eq.(1.23)

$$\eta_{\mu\nu} = \Lambda^\sigma{}_\mu \Lambda^\delta{}_\nu \eta_{\sigma\delta}$$

provides the definition of a Lorentz transformation: it is the set of transformations which leaves the Minkowski inner product invariant, resulting in $\eta = \Lambda^T \eta \Lambda$. Taking the determinant of both sides, we get

$$\det \Lambda = \pm 1. \quad (3.52)$$

If we look at the time component, $\mu = \nu = 0$, then Eq.(1.23) gives

$$\Lambda^0{}_0 = \pm \sqrt{1 + \Lambda^i{}_0 \Lambda^i{}_0}. \quad (3.53)$$

As a consequence, we have four different sub-categories of Lorentz transform:

$$\begin{aligned} L_+^\uparrow &: \det(\Lambda) = +1; & \Lambda^0{}_0 &\geq +1 \\ L_-^\uparrow &: \det(\Lambda) = -1; & \Lambda^0{}_0 &\geq +1 \\ L_+^\downarrow &: \det(\Lambda) = +1; & \Lambda^0{}_0 &\leq -1 \\ L_-^\downarrow &: \det(\Lambda) = -1; & \Lambda^0{}_0 &\leq -1. \end{aligned}$$

Definition 3.15 (Proper orthochronous Lorentz group). *The proper orthochronous Lorentz group (L_+^\uparrow) is the set of Lorentz transformations (Eq.(1.23)) which satisfy $\det(\lambda) = +1$ (proper) and $\Lambda^0{}_0 \geq +1$ (orthochronous). Proper implies that transformations are parity-preserving. Orthochronous implies transformations preserve the direction of time. This group is denoted as $SO(1, 3)_+^\uparrow$, also known as the **restricted Lorentz group**.*

Lorentz transformations not from L_+^\uparrow can be written as combinations of transformations from L_+^\uparrow and the parity (Λ_P) and time-reversal (Λ_T) transformations¹¹

$$\Lambda_P = \text{diag}(1, -1, -1, -1) \quad (3.54)$$

$$\Lambda_T = \text{diag}(-1, 1, 1, 1). \quad (3.55)$$

Hence, we may write the complete Lorentz group as

$$O(1, 3) = \{L_+^\uparrow, \Lambda_P L_+^\uparrow, \Lambda_T L_+^\uparrow, \Lambda_P \Lambda_T L_+^\uparrow\}. \quad (3.56)$$

Notice that only L_+^\uparrow can be built up by infinitesimal transformations from the Lorentz group, because only these transformations are continuously connected to the identity element of the group. There exists a “gap” between the transformations in L_+^\uparrow and those not in L_+^\uparrow , mediated by Λ_P and Λ_T (see Fig. 2). Hence, in what follows, we will concentrate on the Lie group L_+^\uparrow , since the covering group is a single topologically connected piece – whereas the full Lorentz group consists of four topologically separated pieces.

¹¹Note, however, that Λ_P and Λ_T will look quite different in different representations.

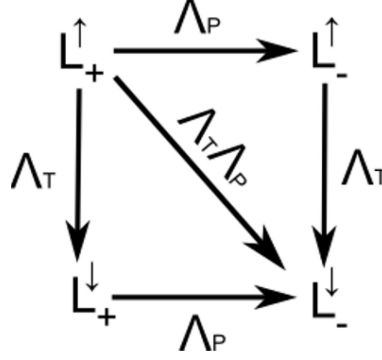


Figure 2. Components of the Lorentz group. Elements of the Lorentz group outside of the proper orthochronous Lorentz group are connected via the discrete “jumps” Λ_P and Λ_T .

3.10. Representation of the Lorentz group

A representation of the Lorentz group can be broken down into 6 generators: 3 generators for **rotations** and 3 generators for **boosts**.

Rotations in space leave the time of an event unchanged. Such transformations obey the characteristic equation of Lorentz transforms Eq.(1.23). We can write down the 3 generators of the rotations J_i as

$$J_i = \begin{pmatrix} 0 & \\ & J_i^{3dim} \end{pmatrix} \quad (3.57)$$

where J_i^{3dim} obey the commutation relations in Eq.(3.17). It follows that J_i also obey these commutation relations.

For the 3 generators of boosts, K_i , consider a series expansion of Λ^μ_ρ , and take the first linear term for an infinitesimal Lorentz transformation

$$\Lambda^\mu_\rho \approx \delta^\mu_\rho + \epsilon K^\mu_\rho + \dots \quad (3.58)$$

Substituting this into Eq.(1.23), we end up with

$$\eta_{\sigma\nu} K^\sigma_\mu + \eta_{\mu\delta} K^\delta_\nu = 0 \quad (3.59)$$

$$K^\sigma_\mu \eta_{\sigma\nu} + \eta_{\mu\delta} K^\delta_\nu = 0 \quad (3.60)$$

Noting that $C^i_j = A^i_k A^k_j$ and using the definition of the transpose (see e.g. Eq.(1.13)),

$$(K^T)^\mu_\sigma \eta_{\sigma\nu} + \eta_{\mu\delta} K^\delta_\nu = 0 \quad (3.61)$$

$$(\mathbf{K}^T \boldsymbol{\eta})_{\mu\nu} + (\boldsymbol{\eta} \mathbf{K})_{\mu\nu} = 0 \quad (3.62)$$

$$\mathbf{K}^T \boldsymbol{\eta} = -\boldsymbol{\eta} \mathbf{K}. \quad (3.63)$$

The transformations generated by K are called **boosts**: namely a change into a coordinate system that moves with a different velocity with respect to the original coordinate system. Now we know, physically, that we can boost into frames which are moving relative to the original frame in some combination of the x , y , and z axes. A boost in the x -direction only should leave the location of events in the y and z directions unchanged. From this information, we can derive a basis for the generators K from Eq.(3.63)

$$K_1 = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad K_2 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad K_3 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}. \quad (3.64)$$

Exponentiation of these generators through $\Lambda = e^{i\phi K}$ yield

$$\Lambda_1 = \begin{pmatrix} \cosh(\phi) & -\sinh(\phi) & 0 & 0 \\ -\sinh(\phi) & \cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \Lambda_2 = \begin{pmatrix} \cosh(\phi) & 0 & -\sinh(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh(\phi) & 0 & \cosh(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Lambda_3 = \begin{pmatrix} \cosh(\phi) & 0 & 0 & -\sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix}. \quad (3.65)$$

We can derive the other components of the Lorentz group by operating the parity and time reversal operators Eq.(3.55) on each index of the generators. We find

$$K_i \xrightarrow{P} -K_i \quad (3.66)$$

$$K_i \xrightarrow{T} -K_i. \quad (3.67)$$

A similar exercise can be done for transformations involving only rotations in space, and we find that the generators J of these transformations are invariant under parity and time reversal.

3.11. Lie algebra of the proper orthochronous Lorentz group

From the explicit forms Eq.(3.57) and Eq.(3.57) we can derive the Lie algebra of L_+^\uparrow by brute force, to give

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.68)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.69)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k. \quad (3.70)$$

Eq.(3.68) shows us that rotation generators are closed under commutation, whereas none of the other combinations are. Note also that a general proper orthochronous Lorentz transformation is of the form

$$\Lambda = e^{i\mathbf{J}\cdot\boldsymbol{\theta} + i\mathbf{K}\cdot\boldsymbol{\phi}}. \quad (3.71)$$

We now complexify the Lie algebra, changing the basis into components of the Lie algebra's two "ideals" N^+ and N^- :

$$N_i^\pm = \frac{J_i \pm iK_i}{2}. \quad (3.72)$$

These generators are now separately closed under commutation, and also commute with each other

$$[N_i^+, N_j^+] = i\epsilon_{ijk}N_k^+ \quad (3.73)$$

$$[N_i^-, N_j^-] = i\epsilon_{ijk}N_k^- \quad (3.74)$$

$$[N_i^+, N_j^-] = 0. \quad (3.75)$$

Notice that the first two commutation relations are precisely the commutation relations we derived for $SU(2)$ in Eq.(3.27)! We have therefore discovered that the complexified Lie algebra of L_+^\uparrow consists of two copies of the Lie algebra of the complexification of $\mathfrak{su}(2)$. Technically, we have shown that

$$\mathfrak{so}(1, 3)_\mathbb{C} \cong \mathfrak{sl}(2, \mathbb{C}) \oplus \mathfrak{sl}(2, \mathbb{C}) \quad (3.76)$$

where $\mathfrak{sl}(2, \mathbb{C})$ is the complexification of $\mathfrak{su}(2)$, \cong denotes an isomorphism, and \oplus denotes the direct sum of a Lie algebra (see Defn. 3.16).

Definition 3.16 (Direct sum of two Lie algebras). For two Lie algebras \mathfrak{g} and \mathfrak{g}' , their direct sum Lie algebra is the vector space $\mathfrak{g} \oplus \mathfrak{g}'$ consisting of all pairs (x, x') , $x \in \mathfrak{g}$, $x' \in \mathfrak{g}'$, with the operation

$$[(x, x'), (y, y')] = ([x, y], [x', y']) \quad (3.77)$$

such that copies of \mathfrak{g} and \mathfrak{g}' commute with each other

$$[(x, 0), (0, x')] = 0. \quad (3.78)$$

In Section 3.8.1 we found that we can label each irreducible representation of $\mathfrak{su}(2)$ by the scalar value of the Casimir element (see Eq.(3.43), we will simply use j instead of $j(j+1)$). Since the Lorentz group consists of two copies of $\mathfrak{su}(2)$, we need two labels to label each irreducible representation of the Lorentz group: j_1 and j_2 where $j_1, j_2 = 0, 1/2, 1, 3/2, \dots$

Finally, note that it is conventional to write the Lie algebra of the Lorentz group more compactly as

$$J_i = \frac{1}{2} \epsilon_{ijk} M_{jk} \quad (3.79)$$

$$K_i = M_{0i} \quad (3.80)$$

where

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho} M_{\nu\sigma} - \eta_{\mu\sigma} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\sigma} + \eta_{\nu\sigma} M_{\mu\rho}). \quad (3.81)$$

3.12. Finite dimensional irreducible representations of the Lorentz group

We denote the finite dimensional irreducible representations of the Lorentz group as (j_1, j_2) where j_1 corresponds to the Casimir element of the N^+ copy of $SU(2)$, and j_2 corresponds to the Casimir element of the N^- copy in Eq.(3.72).

3.12.1. The $(0, 0)$ representation

This is the lowest dimensional representation. Both copies of $SU(2)$ are 1-dimensional, and the only 1×1 matrices that fulfils the commutation relations is the number 0. Hence

$$N_i^+ = N_i^- = 0 \quad (3.82)$$

corresponding to the identity. The $(0, 0)$ representation therefore acts on objects, called **Lorentz scalars**, that do not change under Lorentz transformations. This is called the **scalar representation**. Examples of Lorentz scalars turn out to be spacetime distance between two events; the length of 4-velocities; and the Ricci curvature in a point in spacetime (from General relativity).

3.12.2. The $(\frac{1}{2}, 0)$ representation

In this representation, we use the 2-dimensional representation for one copy of $SU(2)$ $N_i^+ = \sigma_i/2$ (see Eq.(3.26)), and the 1-dimensional representation for the other $N_i^- = 0$. Using Eq.(3.72) we find

$$J_i = \frac{\sigma_i}{2} \quad (3.83)$$

$$K_i = \frac{-i}{2} \sigma_i. \quad (3.84)$$

Letting rotations be denoted by $R_i(\theta)$ and boosts by $B_i(\phi)$, we have

$$R_i(\theta) = e^{\frac{i\theta}{2} \sigma_i} \quad (3.85)$$

$$B_i(\phi) = e^{\frac{\phi}{2} \sigma_i}. \quad (3.86)$$

So, for example, writing out explicitly the form of $R_1(\theta) \equiv R_x(\theta)$, we have

$$R_x(\theta) = \begin{pmatrix} \cos(\theta/2) & i \sin(\theta/2) \\ i \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}. \quad (3.87)$$

Contrast this to the 2D rotation matrix in Eq.(2.2): we see that **we must rotate this object by 4π to get an identity transform!** It is clear that the object which such transformations operate on **are not** simple two-dimensional vectors, or indeed even 4-vectors of Minkowski space. The **two-component** object this representation acts upon are called **left-chiral spinors**.

Definition 3.17 (Left-chiral spinor). *Is an object that transforms under the $(\frac{1}{2}, 0)$ representation of the Lorentz group. We denote them as*

$$\chi_L = \begin{pmatrix} (\chi_L)_1 \\ (\chi_L)_2 \end{pmatrix}. \quad (3.88)$$

We will see later that the two components of this object correspond to **spin-up** and **spin-down**.

Notice how the factor of $\theta/2$ in Eq.(3.87) means that rotation of a spinor by 2π gives a minus sign. A spinor must be rotated by 4π to get the identity transform. We saw from the previous section that the lowest-dimensional representation of the Lorentz group is always the identity, and therefore trivial. Hence, we may say that **a spinor is the simplest object that can be Lorentz-transformed** (Steane, 2013).

3.12.3. The $(0, \frac{1}{2})$ representation

This follows identically to the previous section, except with $N_i^+ = 0$ and $N_i^- = \sigma_i/2$.

$$J_i = \frac{\sigma_i}{2} \quad (3.89)$$

$$K_i = \frac{i}{2} \sigma_i. \quad (3.90)$$

So,

$$R_i(\theta) = e^{\frac{i\theta}{2} \sigma_i} \quad (3.91)$$

$$B_i(\phi) = e^{-\frac{\phi}{2} \sigma_i}. \quad (3.92)$$

We see rotations under this representation are identical to the $(\frac{1}{2}, 0)$ representation, but boosts differ by a minus sign in the exponent. Hence, objects which transform under the $(0, \frac{1}{2})$ representation are **similar but not identical to** left-chiral spinors.

Definition 3.18 (Right-chiral spinor). *Is an object that transforms under the $(0, \frac{1}{2})$ representation of the Lorentz group. We denote them as*

$$\chi_R = \begin{pmatrix} (\chi_R)^1 \\ (\chi_R)^2 \end{pmatrix}. \quad (3.93)$$

The generic name for left- and right-chiral spinors is **Weyl spinors**.

3.12.4. Van der Waerden Notation

Van der Waerden Notation introduces dotted indices to distinguish chirality.

Definition 3.19 (Undotted indices (chiral indices)). *Define a left-chiral spinor with a lower, undotted, index*

$$\chi_L = \chi_a \quad (3.94)$$

Definition 3.20 (Dotted indices (anti-chiral indices)). Define a right-chiral spinor with an upper, dotted, index

$$\chi_R = \chi^{\dot{a}} \quad (3.95)$$

Definition 3.21 (Spinor metric). The spinor metric is defined as

$$\epsilon^{ab} = i(\sigma_y)^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (3.96)$$

and with lower indices

$$\epsilon_{ab} = (\epsilon^{ab})^T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.97)$$

Using the manner in which left and right chiral spinors Lorentz transform, we can show that

$$\chi^{\dot{a}} = \epsilon^{ab} \chi_a^* \quad (3.98)$$

where $*$ denotes complex conjugation. We therefore have the rule that complex conjugation transforms an undotted index into a dotted index, and vice versa:

$$(\chi^{\dot{a}})^* = \chi^a \quad (3.99)$$

$$\chi_a^* = \chi_{\dot{a}}. \quad (3.100)$$

We also find, using the fact that the Pauli matrices are Hermitian ($\sigma^\dagger = \sigma$), that the dot product of spinors of the same type are Lorentz invariant. E.g.

$$\chi_a'^T \chi'^a = \chi_a^T \chi^a = \chi_a^T \epsilon^{ab} \chi_b \quad (3.101)$$

$$\chi_{\dot{a}}'^T \chi'^{\dot{a}} = \chi_{\dot{a}}^T \chi^{\dot{a}} = \chi_{\dot{a}}^T \epsilon^{ab} \chi_b. \quad (3.102)$$

We can thus define the following notation for Lorentz transforming Weyl spinors:

$$\Lambda_{(\frac{1}{2}, 0)} = e^{i\frac{\theta}{2}\sigma + \frac{\phi}{2}\sigma} = \Lambda_a^b \quad (3.103)$$

$$\Lambda_{(0, \frac{1}{2})} = e^{i\frac{\theta}{2}\sigma - \frac{\phi}{2}\sigma} = \Lambda_{\dot{a}}^{\dot{b}}. \quad (3.104)$$

[TODO: I do not understand i) the significance (if any) of the left/right position of the indices in this notation; ii) If the spinor metric also raises/lowers indices of the Lorentz transform. Proceed with caution!] So,

$$\chi_R' = \chi'^{\dot{a}} = \Lambda_{\dot{a}}^{\dot{b}} \chi^{\dot{b}} \quad (3.105)$$

$$\chi_L' = \chi'_a = \Lambda_a^b \chi_b. \quad (3.106)$$

$$(3.107)$$

3.12.5. The $(\frac{1}{2}, \frac{1}{2})$ representation

Let's denote the object which transforms under this representation as $v_a^{\dot{b}}$ using Van der Waerden notation. Since the two copies of $SU(2)$ do not commute within the representation (we have a direct sum), they will not interfere with each other and so we can transform the indices separately. We know that, since we have two copies of $\mathfrak{su}(2)$ with $j = 1/2$, that $v_a^{\dot{b}}$ must have four elements, since each copy of $\mathfrak{su}(2)$ must have two associated states (corresponding to $m = \pm 1/2$, see Section 3.8.1).

We will assume¹² that the irreducible representation acts on 2×2 Hermitian matrices. This is because 2D Hermitian matrices only have 4 degrees of freedom, which coincides with the number of elements we

¹²The justification for the assumption by Schwichtenberg was unclear and seemed circular to me, so I omit it.

argued that $v_a^{\dot{b}}$ must have. Any 2D Hermitian matrix may be decomposed in terms of Pauli matrices which form a basis together with the identity:

$$a_0 1 + a_i \sigma_i. \quad (3.108)$$

For convenience, we first work with v_{ab} , and can use the spinor metric to raise the right index later. By decomposing v_{ab} in terms of Pauli matrices, we have

$$v_{ab} = \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix}. \quad (3.109)$$

We now observe how this object Lorentz transforms

$$v'_{ab} = \Lambda_a^c v_{cd} \Lambda_b^d \quad (3.110)$$

corresponding to (using Eq.(3.104))

$$v \rightarrow v' = v'_{ab} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^d_b \quad (3.111)$$

[TODO: Apparently this is Hermitian but I'm having difficulty proving it... The i in the exponent makes it look very not Hermitian to me...] It turns out, after doing this transformation, that you get exactly the same components as a boosted 4-vector. Therefore the $(\frac{1}{2}, \frac{1}{2})$ representation of $SU(2)$ is exactly the same thing as $SO(1, 3)$. Furthermore, **we see that 4-vectors are not fundamental**, and thus there are physical systems which cannot be described by 4-vectors. In this sense, people sometimes call spinors the square root of 4-vectors: in the same way that vectors are square roots of rank-2 tensors.

3.12.6. Dirac and Majorana spinors

We know from Eq.(3.66) that boosts obtain a sign change under parity transformations. Furthermore, by definition of the generators of the Lorentz group Eq.(3.72) that parity transformations therefore flip $N^+ \leftrightarrow N^-$. This is the motivation for calling the two kinds of Weyl spinor left- and right-handed.

Definition 3.22 (Dirac spinor). *A Dirac spinor is a concatenation of a left-chiral Weyl spinor with an independent right-chiral Weyl spinor*

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \xi^{\dot{a}} \end{pmatrix} \quad (3.112)$$

where the independence of the two components is denoted by the different symbols χ and ξ .

Dirac spinors transform simply by constructing a block-diagonal matrix of transformations for the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ representations, since the two constituent Weyl spinors transform independently:

$$\Psi \rightarrow \Psi' = \Lambda_{(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})} \Psi = \begin{pmatrix} \Lambda_{(\frac{1}{2}, 0)} & 0 \\ 0 & \Lambda_{(0, \frac{1}{2})} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \quad (3.113)$$

Definition 3.23 (Majorana spinor). *A Majorana spinor is a special case of a Dirac spinor*

$$\Psi = \begin{pmatrix} \chi_L \\ \chi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \chi^{\dot{a}} \end{pmatrix}. \quad (3.114)$$

If a transformation which operates on a Dirac spinor is invariant to parity, then we will be able to say that that transformation is invariant to parity. Hence, it is useful to observe how Dirac spinors transform under parity. Since parity transforms result in $N^+ \leftrightarrow N^-$, then

$$\left(\frac{1}{2}, 0\right) \xleftrightarrow{P} \left(0, \frac{1}{2}\right). \quad (3.115)$$

So, instead of transforming under the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation, a parity transformed Dirac spinor transforms under the $(0, \frac{1}{2}) \oplus (\frac{1}{2}, 0)$ representation

$$\Psi^P \rightarrow (\Psi^P)' = \Lambda_{(0, \frac{1}{2}) \oplus (\frac{1}{2}, 0)} \Psi^P = \begin{pmatrix} \Lambda_{(0, \frac{1}{2})} & 0 \\ 0 & \Lambda_{(\frac{1}{2}, 0)} \end{pmatrix} \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \quad (3.116)$$

Hence, a parity transformation has the following effect on a Dirac spinor

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \xrightarrow{P} \Psi^P = \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \quad (3.117)$$

Note that this is **not** the same as $\chi_L \rightarrow \chi_R$, which we discuss next.

3.12.7. Spinor charge conjugation

Definition 3.24. Given a Dirac spinor Ψ

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \quad (3.118)$$

we define the charge conjugate spinor as

$$\Psi^C = \begin{pmatrix} \xi_L \\ \chi_R \end{pmatrix}. \quad (3.119)$$

Note that a charge conjugate spinor still Lorentz transforms like a Dirac spinor, according to $\Lambda_{(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})}$. We show later that charge conjugation flips **all** labels used to describe fundamental particles (not just electric charge). We can see, for example, that a left-chiral spinor would be transformed into a right-chiral spinor under charge conjugation.

3.13. Infinite-dimensional representations of the Lorentz group

So far we've considered transformations on objects Φ_a of the form

$$\Phi_a \rightarrow \Phi'_a = M_{ab}(\Lambda) \Phi_b \quad (3.120)$$

where M_{ab} denotes a particular finite-dimensional representation of the Lorentz transformation Λ . However, objects in physics dynamically change with space and time. Therefore, our objects must become functions of coordinates $\Phi = \Phi(x)$, where the coordinates themselves are also affected by Lorentz transformations. Hence¹³

$$\Phi_a(x) \rightarrow \Phi'_a(x) = M_{ab}(\Lambda) \Phi_b(\Lambda x). \quad (3.121)$$

Our transformations therefore consist of two parts. One part, represented by a finite-dimensional representation, acting on Φ_a ; a second part that transforms the spacetime coordinates. We therefore need an infinite number of basis functions to decompose the object $\Phi_a(x)$ – which is the idea behind Fourier transforms.

¹³Though most authors use the Wigner convention for symmetry operators: $\Phi'_a(x) = M_{ab}(\Lambda) \Phi_b(\Lambda^{-1}x)$

Definition 3.25 (Infinite-dimensional representation of the Lorentz group). *The generator of the infinite-dimensional representation of the Lorentz group is given by differential operators*

$$M_{\mu\nu}^{inf} = i(x^\mu \partial^\nu - x^\nu \partial^\mu) \quad (3.122)$$

Note that we can plug $M_{\mu\nu}^{inf}$ into Eq.(3.81) to show that it does indeed satisfy the definition of the Lorentz algebra.

The complete transformation is then a combination of a finite-dimensional representation $M_{\mu\nu}^{fin}$ and an infinite-dimensional transformation

$$\Phi_a(x) \rightarrow \left(e^{-\frac{i}{2} \omega^{\mu\nu} M_{\mu\nu}^{fin}} \right)_a^b e^{-\frac{i}{2} \omega^{\mu\nu} M_{\mu\nu}^{inf}} \Phi_b(x) \quad (3.123)$$

where the extent of rotation θ_i and boost ϕ_i correspond to

$$\theta_i = \frac{1}{2} \epsilon_{ijk} \omega_{jk} \quad (3.124)$$

$$\phi_i = \omega_{0i} \quad (3.125)$$

(see Eq.(3.79)). Since the matrices $M_{\mu\nu}^{fin}$ are finite dimensional and constant, we can [apparently] put the two components together:

Definition 3.26 (Field representation of the Lorentz group). *The field representation of the Lorentz group is given by:*

$$\Phi_a(x) \rightarrow \left(e^{-\frac{i}{2} \omega^{\mu\nu} M_{\mu\nu}} \right)_a^b \Phi_b(x) \quad (3.126)$$

$$M_{\mu\nu} = M_{\mu\nu}^{fin} + M_{\mu\nu}^{inf}. \quad (3.127)$$

3.14. The Poincaré group

We are now in a position to discuss **translations**. Translations do not mix components, so we do not need a finite-dimensional representation of the group of translations. The Lorentz group plus¹⁴ the group of translations is called the **Poincaré group**. A one-dimensional infinitesimal translation of a function along the x -axis is given by

$$\Phi(x) \rightarrow \Phi(x + \epsilon) = \Phi(x) + \partial_x \Phi(x) \epsilon \quad (3.128)$$

(which is the Taylor expansion of $\Phi(x + \epsilon)$ to first order). In physics, it is convenient to define an i with the generator

$$P_i \equiv -i \partial_i \quad (3.129)$$

$$P_0 \equiv i \partial_0 \quad (3.130)$$

$$\implies P_\mu = i \eta^{\mu\nu} \partial_\nu. \quad (3.131)$$

Hence an arbitrary finite translation of an object by a^i is given by

$$\Phi(x) \rightarrow \Phi(x + a) = e^{-ia^i P_i} \Phi(x) = e^{-a^i \partial_i} \Phi(x). \quad (3.132)$$

Writing this as a Taylor expansion simply yields the Taylor expansion of $\Phi(x + a)$. To translate to another point in time we use P_0 .

¹⁴Technically, the Poincaré group is the semi-direct sum of the Lorentz group and translations.

The Lie algebra of the Poincaré group is

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.133)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.134)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k \quad (3.135)$$

$$[J_i, P_j] = i\epsilon_{ijk}P_k \quad (3.136)$$

$$[J_i, P_0] = 0 \quad (3.137)$$

$$[K_i, P_j] = i\delta_{ij}P_0 \quad (3.138)$$

$$[K_i, P_0] = -iP_i. \quad (3.139)$$

More succinctly, along with Eq.(3.81) we have the additional commutation relations

$$[P_\mu, P_\nu] = 0 \quad (3.140)$$

$$[M_{\mu\nu}, P_\rho] = i(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu) \quad (3.141)$$

It turns out that the Poincaré group has two Casimir operators. The first one is

$$P_\mu P^\mu := m^2. \quad (3.142)$$

We will find out later that m coincides with the **mass** of particles. The second Casimir operator is

$$W_\mu W^\mu \quad (3.143)$$

$$W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}P_\nu M_{\rho\sigma} \quad (3.144)$$

where W^μ is called a **Pauli-Lubanski four-vector**. It can be shown that $j \equiv j_1 + j_2$ can be used as a label (where each j_i corresponds to a representation of $SU(2)$), which we refer to as the **spin** j representation. W^μ describes spin states of moving particles. So, overall, each representation is labelled by two scalar values: m and j . m can take arbitrary values whereas j is restricted to half-integers.

3.15. Elementary particles

The labels for the irreducible representations of the double cover of the Poincaré group, mass m and spin j , are how elementary particles are labelled in physics. More labels, called charges, will follow later from internal symmetries¹⁵: electric charge, weak isospin, and color charge.

The particles of the standard model are one of:

- Spin 0, described by an object Φ , called a **scalar**, that transforms according to the $(0,0)$ representation of the Lorentz group, called the **spin 0 representation**, or scalar representation. The Higgs particle is described by a scalar field.
- Spin $\frac{1}{2}$, described by an object Ψ , called a **spinor**, transforms according to the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation, called the **spin $\frac{1}{2}$ representation**, or spinor representation. Electrons and quarks are described by spinors.
- Spin 1, is described by an object A , called a **vector**, that transforms according to the $(\frac{1}{2}, \frac{1}{2})$ representation, called the **spin 1 representation**, or vector representation. Photons are described by vectors.

¹⁵An internal symmetry is a transformation acting only on the fields, therefore not transforming spacetime points, and leaving the lagrangian or the physical results invariant.

4. The framework: field theory

The underlying idea of field theory is that Nature minimizes *something*. This something must be Lorentz invariant (otherwise we would get different equations of motion for fields for different observers). It will turn out that this *something* is the simplest possible choice of Lorentz invariant.

4.1. The least action principle

Definition 4.1 (Action (classical mechanics)). *The action of a physical system S under classical mechanics is defined as*

$$S[q(t)] = \int_{t_1}^{t_2} \mathcal{L}(q(t), \dot{q}(t), t) dt \quad (4.1)$$

where $q(t)$ is a position vector with some finite number of degrees of freedom, and \mathcal{L} is called the **Lagrangian**.

Note that, in some theories we will go on to consider, we cannot derive a Lorentz-invariant theory by only considering up to the first-order derivative of $q(t)$ and will have to go to second order. **Ostrogradsky instability** prevents us from considering theories with more than two time derivatives, since they correspond to Hamiltonians unbounded from below. It is also interesting to note that a theory with an infinite number of derivatives would allow a non-local theory¹⁶. Furthermore, in order to get a free (non-interacting) theory of particles/fields, the Lagrangian must stop at the lowest-order non-trivial derivative.

Definition 4.2 (Least action principle). *The principle of least action states that the equations of motion correspond to*

$$\delta S = 0 \quad (4.2)$$

where δ means a small change: meaning that the action is stationary to first order for sufficiently short, finite segments in the path.

Definition 4.3 (Lagrangian density). *Define the Lagrangian density \mathcal{L} as*

$$\mathcal{L} = \int \mathcal{L} d^3x \quad (4.3)$$

In **field theories**, we do not use the position/momentum of particles to derive equations of motion, but instead fields $\Phi(\vec{x}, t)$, where the action becomes

Definition 4.4 (Action (field theory)). *The action of a physical system S under a field theory is defined as*

$$S[\Phi(\vec{x}, t)] = \int \mathcal{L}(\Phi(\vec{x}, t), \partial_\mu \Phi(\vec{x}, t), \vec{x}, t) d^4x. \quad (4.4)$$

where $\Phi(\vec{x}, t)$ is a field.

Notice that Eq.(4.4) treats space and time on an equal footing, since fields vary over both space and time. In contrast, for classical mechanics, we parametrized the dynamics of our particle with time only: $q(t)$ and $\dot{q}(t)$.

¹⁶A non-local theory is one which admits interactions of the form $\Phi(x)\Phi(x-h)$ for some arbitrary distance h

4.2. Euler-Lagrange equation

Consider a classical path of a particle between fixed two points in time t_1 and t_2 that obeys the least action principal. We consider perturbing that path, and requiring that the difference in the action to first order vanishes. This yields the Euler-Lagrange equations.

Mathematically, first expand the perturbed Lagrangian around the point (q, \dot{q}, t) using a double Taylor expansion

$$\mathcal{L}(q + \epsilon, \dot{q} + \dot{\epsilon}(t), t) = \mathcal{L}(q, \dot{q}, t) + (q + \epsilon - q) \frac{\partial \mathcal{L}}{\partial t} + (\dot{q} + \dot{\epsilon} - \dot{q}) \frac{\partial \mathcal{L}}{\partial \dot{q}} + \dots \quad (4.5)$$

where $q = q(t)$, $\dot{q} = \dot{q}(t)$, $\epsilon = \epsilon(t)$. By demanding $\delta S = 0$, we have

$$\delta S = \left[\int_{t_1}^{t_2} \mathcal{L}(q, \dot{q}, t) dt \right] - \left[\int_{t_1}^{t_2} \mathcal{L}(q + \epsilon, \dot{q} + \dot{\epsilon}, t) dt \right] = 0 \quad (4.6)$$

So, substituting the first-order Taylor expansion, we have

$$\int_{t_1}^{t_2} \left(\epsilon \frac{\partial \mathcal{L}}{\partial q} + \dot{\epsilon} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) dt = 0. \quad (4.7)$$

We wish to factorize out $\epsilon(t)$ because it is an arbitrary perturbation. Using integration by parts on the term involving $\dot{\epsilon}$

$$\int \dot{\epsilon} \frac{\partial \mathcal{L}}{\partial \dot{q}} dt = \left. \frac{\partial \mathcal{L}}{\partial \dot{q}} \epsilon \right|_{t_1}^{t_2} - \int \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \epsilon dt \quad (4.8)$$

where the first term cancels due to the assumption of fixed boundary conditions, so $\epsilon(t_1) = \epsilon(t_2) = 0$. We thus arrive at

$$\int_{t_1}^{t_2} \epsilon(t) \left[\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \right] dt = 0 \quad (4.9)$$

and therefore we arrive at the famous Euler-Lagrange equation

$$\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial q} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) = 0. \quad (4.10)$$

For (classical) field theory, we follow similar steps to obtain

$$\frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) = 0. \quad (4.11)$$

4.3. Noether's theorem

Noether's theorem shows that each symmetry of the Lagrangian corresponds to a conserved quantity. This is one of the most beautiful insights in the history of science.

In order for dynamics to be invariant, the action must be invariant under a transformation of the Lagrangian. Consider some function $G(q)$ **[TODO: Or is it $G(q, \dot{q})$? Or $G(q, \dot{q}, t)$? Assuming he means what he says, $G(q)$]**. It turns out we can always add the total time derivative of a function to the Lagrangian

$$\mathcal{L} \rightarrow \mathcal{L} + \frac{dG}{dt} \quad (4.12)$$

because

$$\delta S \rightarrow \delta S + \int_{t_1}^{t_2} dt \frac{d}{dt} \delta G = \delta S + \left. \frac{\partial G}{\partial q} \delta q \right|_{t_1}^{t_2} \quad (4.13)$$

due to the vanishing variation δq at the boundaries.

Expanding $\delta\mathcal{L}$, and using the Euler-Lagrange equation Eq.(4.10), and rewriting with the product rule, yields

$$J = \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta q + G \quad (4.14)$$

$$\frac{d}{dt}J = 0 \quad (4.15)$$

And hence $J = \text{const}$, meaning that J is **conserved with time**.

For example, for the Lagrangian $\mathcal{L} = m\dot{q}^2/2$, the Lagrangian is invariant under spatial translation $q \rightarrow q + a$. The corresponding conserved quantity is

$$J_{\text{trans}} = \frac{\partial\mathcal{L}}{\partial\dot{q}}a = m\dot{q}a.$$

$\frac{d}{dt}J = 0$ holds for arbitrary a , and therefore momentum is conserved because the Lagrangian is invariant under spatial translations.

References

Schwichtenberg, J., 2015 *Physics from symmetry*. Springer.

Steane, A. M., 2013 An introduction to spinors. arXiv preprint [arXiv:1312.3824](https://arxiv.org/abs/1312.3824) .