Gregory Naber

Positive Energy Representations of the Poincaré Group

A Sketch of the Positive Mass Case and its Physical Background

December 7, 2016
This is for my grandchildren Amber, Emily, Garrett, and Lacey and my buddy Vinnie.
Preface

During the period in which the foundations of quantum mechanics were being laid it was clear that there were two issues that demanded attention. On the one hand, there were classical physical systems, such as the electromagnetic field, that could not be understood from the perspective of classical mechanics and so were not directly amenable to the techniques that were evolving for the quantization of mechanical systems. On the other hand, at the time this development was taking place the special theory of relativity was a firmly established part of theoretical physics and yet quantum mechanics took no account of it and it was clear that a really satisfactory quantum theory must be “relativistic.” The problem of reconciling quantum mechanics and special relativity is formidable and, indeed, they are essentially irreconcilable if one insists on retaining also the classical notion of a point material particle. Naively, this is due to the fact that special relativity requires that any such particle can be regarded as at rest in some inertial frame of reference and, in such a frame, it would have a well-defined position (the spot where it rests) and momentum (namely, zero) and this would be forbidden by the uncertainty principle of quantum mechanics if, as relativity demands, all such frames of reference must be regarded as physically equivalent.

In a proposed sequel to the manuscript [Nab5] on the Foundations of Quantum Mechanics we will attempt to sort out, for those whose background is in mathematics rather than physics, how such a reconciliation might be carried out and how the notion of a quantum field emerges as a result. The brief manuscript before you now is an excerpt from this sequel that, we hope, may be of some independent interest. It centers around the problem of building the machinery required to define what it means to say that a quantum theory is “relativistic”.

Chapter 1 is a synopsis of the material on Lie groups, Lie algebras and their representations that we will require. Sections 1.1 and 1.2 contain the basic definitions and examples, while Section 1.3 introduces the important notion of a projective representation of a Lie group. In Sections 1.4 and 1.5 we describe the so-called Mackey machine for constructing all of the irreducible, unitary representations of certain semi-direct products of Lie groups. There are some rather deep theorems here that
we will state, but not prove. We will, however, make every effort to provide ample references for all of the details we do not include.

In Chapter 2 we turn our attention special relativity. The first three sections motivate and describe the basic geometrical structure of Minkowski spacetime $\mathcal{M}$ and are drawn largely from [Nab4]. Section 2.4 contains the construction of the 2-fold universal covering groups $\text{SL}(2, \mathbb{C})$ and $\text{ISL}(2, \mathbb{C})$ of the Lorentz and Poincaré groups $\mathcal{L}_+^1$ and $\mathcal{P}_+^1$, respectively. In Section 2.5 we study the structure of the Lie algebras of $\mathcal{L}_+^1$ and $\mathcal{P}_+^1$ (and therefore of $\text{SL}(2, \mathbb{C})$ and $\text{ISL}(2, \mathbb{C})$). Here we also introduce the universal enveloping algebra $\mathfrak{U}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$ since this is where one finds the so-called Casimir invariants. The application of the Mackey machine to $\text{ISL}(2, \mathbb{C})$ requires some rather detailed information about the algebraic dual of Minkowski spacetime, called momentum space, and all of this will be derived in Section 2.6.

In Section 2.7 we identify the relativistic invariance of a quantum system with the existence of a projective representation of $\mathcal{P}_+^1$ on the Hilbert space $\mathcal{H}$ of the system and indicate how the problem of finding all of these can be reduced, via a deep theorem of Bargmann, to that of finding the unitary representations of the double cover $\text{ISL}(2, \mathbb{C})$ of $\mathcal{P}_+^1$. Finally, we pull all of this information together in Section 2.8 to enumerate those particular irreducible, unitary representations of $\text{ISL}(2, \mathbb{C})$ that, from the perspective of the physical applications we have in mind, are of interest to us, namely, those of positive mass and positive energy. We will not consider the massless case, but will supply references.

The source of our interest in the irreducible, projective representations of the Poincaré group resides in physics. Specifically, these are precisely the objects required for a rigorous definition of the relativistic invariance of a quantum theory. One need not understand the physical motivation in order to appreciate the mathematics, but one then knows only half of the story. For those inclined toward full disclosure we have included Appendix A with brief discussions of some physical background material on classical and quantum mechanics, taken largely from [Nab5]. For our purposes here the central notion is that of a symmetry or, more generally, a symmetry group of a physical system. In Sections A.2 and A.3 we will try to describe enough of the formalism of classical Lagrangian and Hamiltonian mechanics to understand what it means to say that a classical mechanical system admits a symmetry group and what can be inferred from this. In Section A.4 we will attempt the same thing for quantum mechanics and this will lead us to the projective representations of $\mathcal{P}_+^1$. The Appendix concludes with Section A.5 in which we provide a brief introduction to the quantum mechanical phenomenon of spin.

A few editorial comments are worth making at the outset. The subtitle we have chosen contains the word “sketch” and this is entirely appropriate. The rather modest goal here is to provide something of a global picture of the ideas, both mathematical and physical, that are involved in understanding the construction and significance of one particular class of irreducible, unitary representations of the Poincaré group and its universal cover. We will not pretend to have offered an exhaustive treatment. We will, however, make a concerted effort to direct those who need the full story to appropriate sources for all of the details we have not included here.
We must also confess that the manuscript assumes a fairly substantial mathematical background. We have provided brief synopses of some of the mathematical machinery required to carry out our task, but generally this has been limited to topics that were not discussed in some detail in [Nab5]. Naturally, it does not matter where this background has been acquired and when it comes time to provide references we will include some readily accessible sources in addition to [Nab5].

We will consistently employ the Einstein summation convention according to which a repeated index, one subscript and one superscript, indicates a sum over the range of values that the index can assume. For example, if \( i \) and \( j \) are indices that range over \( 1 \rightarrow ... \rightarrow n \), then

\[
X^i \frac{\partial L}{\partial \dot{q}^i} = \sum_{i=1}^{n} X^i \frac{\partial L}{\partial \dot{q}^i} = X^1 \frac{\partial L}{\partial \dot{q}^1} + \cdots + X^n \frac{\partial L}{\partial \dot{q}^n}
\]

(a superscript in the denominator counts as a subscript), whereas, if \( \alpha \) and \( \beta \) take the values 0, 1, 2, 3, then

\[
\eta_{\alpha\beta} \nu^\alpha w^\beta = \sum_{\alpha, \beta=0}^{3} \eta_{\alpha\beta} \nu^\alpha w^\beta = \eta_{00} \nu^0 w^0 + \cdots + \eta_{03} \nu^0 w^3 + \cdots + \eta_{30} \nu^3 w^0 + \cdots + \eta_{33} \nu^3 w^3,
\]

and so on.

We should also say a few words about the signature of the quadratic form of Minkowski spacetime. There was a time when the world was evenly divided between \((+++)\) and \((+---)\) and heartfelt arguments were put forth in favor of each. However, there is some statistical evidence that \((+---)\) has won the day. If so, this would put our reference [Nab4] on the losing side. In any case we have opted for \((+---)\) here. This changes nothing essential beyond a few minus signs and the sense of a few inequalities in the definitions. Even so, considering the number of references we will make to [Nab4], we thought it only fair to alert the reader to these adjustments. Forewarned is forearmed.
## Contents

1 Lie Groups and Representations .............................................. 1  
  1.1 Lie Groups .......................................................... 1  
  1.2 Unitary Group Representations ..................................... 14  
  1.3 Projective Representations .......................................... 24  
  1.4 Induced Representations .............................................. 29  
  1.5 Representations of Semi-Direct Products ......................... 33  

2 Minkowski Spacetime ........................................................... 43  
  2.1 Introduction .......................................................... 43  
  2.2 Motivation ............................................................ 43  
  2.3 Geometrical Structure of $\mathcal{M}$ ................................ 48  
  2.4 Lorentz and Poincaré Groups ....................................... 57  
  2.5 Poincaré Algebra ..................................................... 64  
     2.5.1 Introduction .................................................... 64  
     2.5.2 Lie Algebra of $L^+$ ........................................... 65  
     2.5.3 Lie Algebra of $\mathfrak{p}^+$ ................................... 70  
     2.5.4 Universal Enveloping Algebra and Casimir Invariants ....... 85  
  2.6 Momentum Space ....................................................... 92  
     2.6.1 Orbits and Isotropy Groups .................................. 92  
     2.6.2 Invariant Measures ............................................. 99  
     2.6.3 Momentum Space and the Character Group ................. 102  
  2.7 Spacetime Symmetries and Projective Representations .......... 105  
  2.8 Positive Energy Representations of $\mathfrak{p}^+$ with $m > 0$ .... 107  

A Physical Background .......................................................... 123  
  A.1 Introduction .......................................................... 123  
  A.2 Finite-Dimensional Lagrangian Mechanics ......................... 123  
  A.3 Finite-Dimensional Hamiltonian Mechanics ....................... 133  
  A.4 Postulates of Quantum Mechanics .................................. 147  
  A.5 Spin ................................................................. 164
Chapter 1
Lie Groups and Representations

1.1 Lie Groups

We will record here only those specific items that we require in what follows. Essentially everything we need is treated concisely, but in detail in Chapter 3 of [Warn]. One can also consult Chapter 10, Volume I, of [Sp2], Section 5.8 of [Nab2], or, for a much more comprehensive treatment, [Knapp].

A Lie group is a group $G$ that is also a ($C^\infty$) differentiable manifold for which the operations of multiplication $(x,y) \mapsto xy : G \times G \to G$ and inversion $x \mapsto x^{-1} : G \to G$ are smooth ($C^\infty$). If $G_1$ and $G_2$ are Lie groups, then a diffeomorphism $\varphi : G_1 \to G_2$ of $G_1$ onto $G_2$ that is also a group isomorphism is an isomorphism of Lie groups and, if such a thing exists, $G_1$ and $G_2$ are said to be isomorphic. An isomorphism of $G$ onto itself is called an automorphism of $G$ and the set of all such, denoted $\text{Aut}(G)$, is a group under composition, called the automorphism group of $G$.

**Note:** Manifolds are assumed to be Hausdorff and second countable and they are always locally compact so the same is true of Lie groups. In particular, every Lie group $G$ admits a Haar measure, that is, a nonzero Radon measure $\mu_G$ on $G$ that is left-invariant in the sense that $\mu_G(gB) = \mu_G(B)$ for every $g \in G$ and every Borel set $B$ in $G$, where $gB = \{gb : b \in B\}$ (see Section 2.2 of [Fol2]).

The requirement that inversion is a smooth map is actually superfluous. The following is Theorem 5.8.1 of [Nab2].

**Theorem 1.1.1.** Let $G$ be a group that is also a differentiable manifold for which group multiplication $(x,y) \mapsto xy : G \times G \to G$ is smooth. Then $G$ is a Lie group.

Some obvious examples are $\mathbb{R}$ with real number addition, $\mathbb{C}$ with complex addition, $\mathbb{R}^n$ and $\mathbb{C}^n$ with coordinatewise addition, the nonzero real numbers $\mathbb{R}^*$ with real
number multiplication, the nonzero complex numbers \( \mathbb{C}^\times \) with complex multiplication, and the complex numbers of modulus one \( S^1 \) (also denoted \( \mathbb{T} \) in this context) with complex multiplication. The general linear groups \( \text{GL}(n, \mathbb{R}) \) and \( \text{GL}(n, \mathbb{C}) \) consisting of all nonsingular \( n \times n \) matrices with real and complex entries, respectively, are also Lie groups under matrix multiplication. The product of two Lie groups is a Lie group with the product manifold structure and the direct product group structure. We will see more examples shortly. The following is Theorem 3.21 of [Warn].

**Theorem 1.1.2.** (Closed Subgroup Theorem) A closed subgroup \( H \) of a Lie group \( G \) is an embedded submanifold of \( G \) and a Lie group with respect to the induced manifold structure and the group operations inherited from \( G \).

The proof of the following Corollary is a nice application of the Closed Subgroup Theorem so we will include the argument.

**Corollary 1.1.3.** Let \( G \) and \( H \) be Lie groups and \( \phi : G \to H \) a continuous group homomorphism. Then \( \phi \) is smooth.

**Proof.** Let \( \Gamma_\phi : G \to G \times H \) be the graph map defined by

\[
\Gamma_\phi(g) = (g, \phi(g)).
\]

Let \( \pi_G \) and \( \pi_H \) be the projections of \( G \times H \) onto \( G \) and \( H \), respectively. \( \Gamma_\phi \) is continuous, injective, and maps onto the graph \( \Gamma_\phi(G) \) of \( \phi \). Its inverse is the restriction of \( \pi_G \) to \( \Gamma_\phi(G) \) and this is also continuous. Consequently, \( \Gamma_\phi \) is a homeomorphism of \( G \) onto \( \Gamma_\phi(G) \). Since \( \phi \) is a homomorphism, \( \Gamma_\phi(G) \) is a subgroup of the Lie group \( G \times H \). It is closed because it is the inverse image of the diagonal in \( H \times H \) under the continuous map \( G \times H \to H \times H \) defined by \( (g, h) \mapsto (\phi(g), h) \).

According to the Closed Subgroup Theorem, \( \Gamma_\phi(G) \) is an embedded submanifold of the Lie group \( G \times H \) and is itself a Lie group under the group operation on \( G \times H \) and the submanifold structure. Now notice that

\[
\phi = (\pi_H \vert_{\Gamma_\phi(G)}) \circ (\pi_G \vert_{\Gamma_\phi(G)})^{-1}.
\]

The projections \( \pi_G : G \times H \to G \) and \( \pi_H : G \times H \to H \) are smooth and, because \( \Gamma_\phi(G) \) is an embedded submanifold of \( G \times H \), the inclusion map \( \iota : \Gamma_\phi(G) \hookrightarrow G \times H \) is smooth. Since \( \pi_H \vert_{\Gamma_\phi(G)} = \pi_H \circ \iota \), it is also smooth. Similarly, \( \pi_G \vert_{\Gamma_\phi(G)} = \pi_G \circ \iota \) is smooth. To show that \( \phi \) is smooth it will therefore suffice to show that the homeomorphism \( \pi_G \vert_{\Gamma_\phi(G)} \) is a diffeomorphism. Moreover, since \( \Gamma_\phi(G) \) is a Lie group, left translation by any element is a diffeomorphism so it will be enough to show that \( \pi_G \vert_{\Gamma_\phi(G)} \) is a local diffeomorphism near some point. Since \( \pi_G \vert_{\Gamma_\phi(G)} \) is smooth this follows from Sard’s Theorem which asserts that \( \pi_G \vert_{\Gamma_\phi(G)} \) must have a regular value.
Remark 1.1.1. If you would prefer to avoid an application of Sard’s Theorem there is a direct proof of the last assertion in Lemma 9.4 of [VDB].

Example 1.1.1. A closed subgroup of some general linear group $GL(n, \mathbb{R})$ or $GL(n, \mathbb{C})$ is called a matrix Lie group. We enumerate a few of these.

1. The special linear groups $SL(n, \mathbb{R})$ and $SL(n, \mathbb{C})$ consist of all of those elements of $GL(n, \mathbb{R})$ and $GL(n, \mathbb{C})$, respectively, that have determinant one. As manifolds, $SL(n, \mathbb{R})$ and $SL(n, \mathbb{C})$ have dimension $n^2 - 1$. When $n = 2$ the special linear group $SL(2, \mathbb{C})$ is closely related to the Lorentz group (see Section 2.4).

2. The orthogonal group $O(n)$ consists of all $n \times n$ real matrices $A$ that satisfy $AA^T = A^T A = \text{id}_{n \times n}$. These are precisely the matrices which, when identified with linear transformations on $\mathbb{R}^n$, preserve the standard inner product $\langle x, y \rangle = x^1 y^1 + \cdots + x^n y^n$, that is, which satisfy $\langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{R}^n$. These all have determinant $\pm 1$. As a manifold, $O(n)$ has dimension $n(n-1)/2$ and two connected components corresponding to $\det A = 1$ and $\det A = -1$.

3. The special orthogonal group $SO(n)$ is the subgroup of $O(n)$ consisting of those elements that have determinant 1. It is an open subset of $O(n)$ so the dimension of $SO(n)$ is also $n(n-1)/2$ and it is the connected component of $O(n)$ containing the identity. When $n = 3$ we will refer to the special orthogonal group $SO(3)$ as the rotation group. We will see in Theorem A.2.2 below that the elements of $SO(3)$ are the matrices, with respect to orthonormal bases for $\mathbb{R}^3$, of rotations about some axis.

4. The unitary group $U(n)$ consists of all $n \times n$ complex matrices $A$ that satisfy $\overline{A} A^T = A^T A = \text{id}_{n \times n}$, where $\overline{A}^T$ is the conjugate transpose of $A$. These are precisely the matrices which, when identified with linear transformations on $\mathbb{C}^n$, preserves the standard Hermitian inner product $\langle x, y \rangle = x^1 y^1 + \cdots + x^n y^n$, that is, which satisfy $\langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{C}^n$. The determinant of any element of $U(n)$ is a complex number of modulus one. As a manifold, $U(n)$ has dimension $n^2$. Note that we have taken the Hermitian inner product to be complex-linear in the second slot and conjugate linear in the first.

5. The special unitary group $SU(n)$ is the subgroup of $U(n)$ consisting of those elements that have determinant 1. As a manifold, $SU(n)$ has dimension $n^2 - 1$. When $n = 2$ the special unitary group $SU(2)$ is closely related to the rotation group $SO(3)$ (see Section 2.4).

6. If $n$ is a positive integer and $p$ and $q$ are non-negative integers with $n = p+q$, then the semi-orthogonal group $O(p, q)$ consists of all $n \times n$ real matrices $A$ satisfying $A^T \eta A = \eta$, where
These are precisely the matrices which, when identified with linear transformations on \( \mathbb{R}^n \), preserve the standard indefinite inner product

\[
(x, y) = x^1 y^1 + \cdots + x^p y^p - x^{p+1} y^{p+1} - \cdots - x^n y^n
\]
of index \( q \) on \( \mathbb{R}^n \), that is, which satisfy \( \langle Ax, Ay \rangle = \langle x, y \rangle \) for all \( x, y \in \mathbb{R}^n \). These all have determinant \( \pm 1 \). As a manifold, \( O(p, q) \) has dimension \( n(n - 1)/2 \). When \( p = 1 \) and \( q = 3 \), the semi-orthogonal group \( O(1, 3) \) is the general Lorentz group (see Section 2.4). The elements of \( O(1, 3) \) are generally denoted \( A = (\Lambda_{\alpha\beta})_{\alpha, \beta=0,1,2,3} \) and they all satisfy \( \det A = \pm 1 \) and either \( \Lambda_{00} \geq 1 \) or \( \Lambda_{00} \leq -1 \). These four possibilities determine the four connected components of \( O(1, 3) \).

7. The **special semi-orthogonal group** \( \text{SO}(p, q) \) is the subgroup of \( O(p, q) \) consisting of those elements with determinant 1. It is an open subset of \( O(p, q) \) so the dimension of \( \text{SO}(p, q) \) is also \( n(n - 1)/2 \). When \( p = 1 \) and \( q = 3 \), the special semi-orthogonal group \( \text{SO}(1, 3) \) is the proper Lorentz group (see Section 2.4). It is the union of two of the four connected components of \( O(1, 3) \), one of which contains the identity matrix. The component of \( O(1, 3) \) containing the identity is denoted \( \text{SO}^+(1, 3) \) and is just the proper, orthochronous, Lorentz group \( \mathcal{L}_{\text{+}} \) (see Section 2.4).

8. Lie groups that are not given directly as closed subgroups of some \( \text{GL}(n, \mathbb{R}) \) or \( \text{GL}(n, \mathbb{C}) \) can often be shown to be isomorphic to matrix Lie groups. For our purposes the most important examples will arise as “semi-direct products” of the matrix Lie groups described above. Semi-direct products are discussed in general in Section 1.5. There we will illustrate the process of re-interpreting them as matrix Lie groups for the inhomogeneous rotation group \( \text{ISO}(3) \). The most important example, however, is the inhomogeneous Lorentz group, or Poincaré group, which will be described in detail in Section 2.4.

A **real Lie algebra** is a finite-dimensional, real vector space \( \mathcal{A} \) on which is defined a (real) bilinear operation \([ , ] : \mathcal{A} \times \mathcal{A} \to \mathcal{A} \), called **bracket**, which is skew-symmetric

\[
[Y, X] = -[X, Y] \quad \forall X, Y \in \mathcal{A}
\]
and satisfies the **Jacobi identity**

\[
[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0 \quad \forall X, Y, Z \in \mathcal{A}.
\]

A **complex Lie algebra** is defined in exactly the same way, but with \( \mathcal{A} \) a complex vector space and \([ , ] \) complex-bilinear.
Remark 1.1.2. Any real Lie algebra $A$ has a Lie algebra complexification $A_C$ defined in the following way. The points of $A_C$ are ordered pairs $(X_1, X_2)$ of vectors in $A$ and one defines addition and complex scalar multiplication in $A_C$ by $(X_1, X_2) + (Y_1, Y_2) = (X_1 + Y_1, X_2 + Y_2)$ and $(a + bi)(X_1, X_2) = (aX_1 - bX_2, aX_2 + bX_1)$. This gives $A_C$ the structure of a complex vector space. It is customary to write $(X_1, X_2)$ as $X_1 + iX_2$ so that these operations take the same form as those of $C$, that is,

$$(X_1 + iX_2) + (Y_1 + iY_2) = (X_1 + Y_1) + i(X_2 + Y_2)$$

and

$$(a + bi)(X_1 + iX_2) = (aX_1 - bX_2) + i(bX_1 + aX_2).$$

The bracket $[,]_C$ on $A_C$ is then defined by just “multiplying out”, that is,

$$[X_1 + iX_2, Y_1 + iY_2]_C = ([X_1, Y_1] - [X_2, Y_2]) + i([X_1, Y_2] + [X_2, Y_1]).$$

If $A_1$ and $A_2$ are two Lie algebras with brackets $[,]_1$ and $[,]_2$, respectively, then a linear map $T : A_1 \to A_2$ that satisfies $T([X, Y]_1) = [T(X), T(Y)]_2$ for all $X, Y \in A_1$ is called a Lie algebra homomorphism. If $A_2$ is the algebra $End(V)$ of endomorphisms of some complex vector space $V$ with the commutator bracket, then $T$ is called a Lie algebra representation of $A_1$. If $T : A_1 \to A_2$ is a linear isomorphism then it is called a Lie algebra isomorphism and, if such a thing exists, we say that $A_1$ and $A_2$ are isomorphic. A Lie algebra isomorphism of $A$ onto itself is called an automorphism of $A$. A Lie subalgebra of a Lie algebra $A$ is a linear subspace $B$ of $A$ that is closed under the bracket $[,]$ of $A$ and therefore is a Lie algebra in its own right with this same bracket.

Every Lie group $G$ has associated with it a Lie algebra $g$ that can be defined in two equivalent ways. A vector field $X$ on $G$ is said to be left invariant if, for each $g \in G$, $(L_g)_o X = X \circ L_g$, where $L_g : G \to G$ is the left translation diffeomorphism $L_g(g') = gg'$ for every $g' \in G$ and $(L_g)_o$ is its derivative. Left invariant vector fields are necessarily smooth (Theorem 5.8.2 of [Nab2]). Every tangent vector at the identity in $G$ gives rise, via left translation, to a unique left-invariant vector field on $G$. One can think of $g$ as the real linear space of left invariant vector fields on $G$ with $[X, Y]$ taken to be the Lie bracket of the vector fields $X$ and $Y$ (see pages 263-264 of [Nab2] for Lie brackets). Equivalently, $g$ can be identified with the tangent space $T_g(G)$ at the identity $e$ in $G$ with the bracket of two tangent vectors $x$ and $y$ in $T_e(G)$ defined by writing $x = X(e)$ and $y = Y(e)$ for left invariant vector fields $X$ and $Y$ and setting $[x, y] = [X, Y](e)$. In particular, the linear dimension of $g$ is the same as the manifold dimension of $G$.

Exercise 1.1.1. Work all of this out for the additive group $\mathbb{R}^n$. Specifically, prove each of the following.

1. Vector addition is a smooth map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$ so $\mathbb{R}^n$ is a Lie group.
2. The tangent space at any point in \( \mathbb{R}^n \) can be canonically identified with \( \mathbb{R}^n \).
3. If \( L_\alpha \) is the left translation map \( L_\alpha(b) = a + b \) on \( \mathbb{R}^n \) and tangent spaces are identified with \( \mathbb{R}^n \), then the derivative \( (L_\alpha)_* \) is the identity at each point.
4. The left-invariant vector fields on \( \mathbb{R}^n \) are precisely the constant vector fields and the Lie bracket of any two left-invariant vector fields is the zero vector field. Conclude that the Lie algebra of \( \mathbb{R}^n \) is isomorphic to \( \mathbb{R}^n \) with the trivial bracket \([\cdot, \cdot] \equiv 0\) and therefore to \( T_0(\mathbb{R}^n) \) with the trivial bracket.

The general linear group \( G = \text{GL}(n, \mathbb{R}) \) is an open submanifold of \( \mathbb{R}^{n^2} \) so its tangent space at the identity matrix is linearly isomorphic to \( \mathbb{R}^{n^2} \), that is, to the space of all real \( n \times n \) matrices. Thought of in this way the bracket in the Lie algebra \( \mathfrak{gl}(n, \mathbb{R}) \) of \( \text{GL}(n, \mathbb{R}) \) is just the matrix commutator (see pages 278-279 of [Nab2]). More generally, the Lie algebra of any matrix Lie group is some set of real matrices with the bracket given by matrix commutator (see page 279 of [Nab2]). Thus, to determine the Lie algebra of a matrix Lie group \( G \) one need only identify it with a set of matrices and this one can do by computing velocity vectors to smooth curves in \( G \) through \( e \). We will work out one simple example here and then just list some more. The example of real interest to us is in Section 2.5.

**Example 1.1.2.** We will identify the Lie algebra \( \mathfrak{o}(n) \) of the orthogonal group \( O(n) \). Any element of the tangent space \( T_{id}(O(n)) \) is \( A'(0) \) for some smooth curve \( A : (-\epsilon, \epsilon) \to O(n) \) in \( O(n) \) with \( A'(0) = \text{id} \). Since \( O(n) \) is a submanifold of \( \mathbb{R}^{n^2} \) we can regard \( A \) as a smooth curve in \( \mathbb{R}^{n^2} \) and use standard coordinates to differentiate entrywise. Denote the entries of \( A \) by \( A^i_j, i, j = 1, \ldots, n \), and the standard coordinates on \( \mathbb{R}^{n^2} \) by \( x^i_j, i, j = 1, \ldots, n \). Thus, the components of \( A'(0) \) relative to \( \frac{\partial}{\partial x^i_j} |_{id} \) are \( (A^i_j)'(0) \). Since \( A(t) \in O(n) \) for each \( t \in (-\epsilon, \epsilon) \), \( A(t)A(t)^T = \text{id} \), that is,

\[
\sum_{k=1}^{n} A^k_i(t)A^k_j(t) = \delta^{ij}
\]

for each \( t \). Differentiating at \( t = 0 \) gives

\[
\sum_{k=1}^{n} [(A^k_i)'(0)\delta^{jk} + \delta^{ik}(A^k_j)'(0)] = 0
\]

\[
(A^i_j)'(0) + (A^j_i)'(0) = 0
\]

\[
(A^i_j)'(0) = -(A^j_i)'(0)
\]

so \( (A^i_j)'(0) \) is a real, \( n \times n \), skew-symmetric matrix. Thus, \( T_{id}(O(n)) \) is contained in the linear subspace of \( \mathfrak{gl}(n, \mathbb{R}) \) consisting of the skew-symmetric matrices. But the dimension of this subspace is \( n(n-1)/2 \) and this is precisely the dimension of \( O(n) \) so \( \mathfrak{o}(n) \) is precisely the set of real, \( n \times n \), skew-symmetric matrices under commutator.

Notice that, since \( S\text{O}(n) \) is an open submanifold of \( O(n) \), its tangent space at the identity is the same as that of \( O(n) \) so
1.1 Lie Groups

\[ \mathfrak{so}(n) = \mathfrak{o}(n). \]

\( \mathfrak{o}(n) \) and \( \mathfrak{SO}(n) \) are not isomorphic as Lie groups, but they have the same Lie algebras. Two Lie groups are said to be \textit{locally isomorphic} if their Lie algebras are isomorphic.

Once the Lie algebra \( \mathfrak{g} \) of a matrix Lie group \( G \) is identified with a particular vector space of matrices it is often convenient to have in hand a more explicit description of the commutator bracket. For this one can select a basis \( \{ X_i \}_{i=1,...,n} \) for \( \mathfrak{g} \). The bracket on \( \mathfrak{g} \) is then completely determined by the commutators \( [X_i, X_j] \) for \( i, j = 1, \ldots, n \). But each \( [X_i, X_j] \) is a linear combination of the basis elements so there exist constants \( C_{ijk}, k = 1, \ldots, n \), such that

\[ [X_i, X_j] = \sum_{k=1}^{n} C_{ijk} X_k. \]

These are called the \textit{structure constants} of \( \mathfrak{g} \) relative to the chosen basis and they determine the bracket completely.

\textbf{Exercise 1.1.2.} Show that

\[ X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \]

is a basis for \( \mathfrak{o}(3) \) (and \( \mathfrak{so}(3) \)) and compute the commutators to show that

\[ [X_i, X_j] = \epsilon_{ijk} X_k, \quad i, j = 1, 2, 3, \]

where \( \epsilon_{ijk} \) is the Levi-Civita symbol (1 if \( ijk \) is an even permutation of 123, -1 if \( ijk \) is an odd permutation of 123, and 0 otherwise).

\textbf{Example 1.1.3.} It is not a trivial matter to come up with Lie groups that are \textit{not} isomorphic to matrix Lie groups and all of the examples of interest to us are, in fact, isomorphic to matrix Lie groups (although perhaps not obviously so). For each of the following matrix Lie groups we will specify the set of matrices that constitute its Lie algebra; the bracket is always matrix commutator.

1. The Lie algebra \( \mathfrak{gl}(n, \mathbb{R}) \) of \( \text{GL}(n, \mathbb{R}) \) is the linear space of all real, \( n \times n \) matrices.

2. The Lie algebra \( \mathfrak{gl}(n, \mathbb{C}) \) of \( \text{GL}(n, \mathbb{C}) \) is the \textit{real} linear space of all \textit{complex}, \( n \times n \) matrices.

3. The Lie algebras \( \mathfrak{o}(n) \) and \( \mathfrak{so}(n) \) of \( \mathfrak{O}(n) \) and \( \mathfrak{SO}(n) \), respectively, both consist of all real, \( n \times n \), skew-symmetric matrices.
4. The Lie algebra $u(n)$ of $U(n)$ is the real linear space of all complex, $n \times n$ matrices $X$ that are skew-Hermitian ($X^T = -X$).

5. The Lie algebra $su(n)$ of $SU(n)$ is the real linear space of all complex, $n \times n$ matrices $X$ that are skew-Hermitian ($X^T = -X$) and tracefree (Trace $(X) = 0$).

Remark 1.1.3. In Exercise 1.2.6 (9) we will exhibit a basis for $su(2)$ in terms of the Pauli spin matrices and you will use it to show that $su(2)$ is isomorphic to $so(3)$ so that $SU(2)$ and $SO(3)$ are locally isomorphic.

6. The Lie algebras $o(p \hookrightarrow q)$ and $so(p \hookrightarrow q)$ of $O(p \hookrightarrow q)$ and $SO(p \hookrightarrow q)$, respectively, both consist of all real, $n \times n$ matrices $X$ that satisfy $X^T = -\eta X \eta$, where

$$\eta = \text{diag}(1, \ldots, 1, -1, \ldots, -1).$$

When $p = 1$ and $q = 3$ this implies that $o(1, 3) = so(1, 3)$ consists of all real matrices of the form

$$X = \begin{pmatrix} 0 & X_0^0 & X_0^1 & X_0^2 & X_0^3 \\ X_1^0 & 0 & -X_1^1 & X_1^2 & X_1^3 \\ X_2^0 & X_2^1 & 0 & -X_2^2 & X_2^3 \\ X_3^0 & -X_3^1 & X_3^2 & 0 & \end{pmatrix}. $$

This is also the Lie algebra $so^*(1, 3)$ of the proper, orthochronous Lorentz group $SO^+(1, 3) = \mathcal{L}^+_1$ (see Section 2.4).

7. For Lie groups such as the Poincaré group (see Section 2.4) which are given as semi-direct products one can define the general notion of a semi-direct product of Lie algebras and prove that the Lie algebra of a semi-direct product of Lie groups is the semi-direct product of the Lie algebras of these groups (see pages 301-306 of [Nab5] for a brief description and an example or Section I.4 of [Knapp] for the details). One can often avoid this abstract description of the Lie algebra by finding an explicit representation of the Lie group semi-direct product as a matrix Lie group; such a representation of the Poincaré group is described in Section 2.4.

Example 1.1.4. An associative algebra over a field $K$ is a vector space $A$ over $K$ on which is defined a $K$-bilinear map $B : A \times A \rightarrow A$, written simply $B(a, b) = ab$ for all $a, b \in A$ and called multiplication, that satisfies $a(bc) = (ab)c$ for all $a, b, c \in A$. $A$ is said to be unital if there exists an element $1 \in A$, called the unit, satisfying $1a = a1 = a$ for all $a \in A$. A subalgebra of $A$ is a linear subspace $B$ of $A$ that is closed under multiplication and is therefore an algebra in its own right with the same operations as $A$. If $A$ is unital, then $B$ is also required to contain the unit 1. On the other hand, a (two-sided) ideal in $A$ is an additive subgroup $J$ of $A$ with the property that $x \in J$ implies $ax \in J$ and $xa \in J$ for all $a \in A$. If $A$ is unital and $J$
contains the unit, then, in fact, $\mathfrak{j} = \mathcal{A}$. If $S$ is an arbitrary subset of $\mathcal{A}$, then the \textit{ideal generated by $S$} is the intersection of all ideals in $\mathcal{A}$ that contain $S$.

Any associative algebra $\mathcal{A}$ over $\mathbb{R}$ or $\mathbb{C}$ can be given the structure of a Lie algebra by defining on it the \textit{commutator bracket}

$$[a, b] = ab - ba.$$  

We call this the \textit{commutator Lie algebra structure} of $\mathcal{A}$. In Section 2.5.4 we will discuss the problem of representing a given Lie algebra as the commutator Lie algebra of some associative algebra.

If $G$ is a Lie group with Lie algebra $\mathfrak{g}$ we define the \textit{exponential map}

$$\exp : \mathfrak{g} \to G$$

from $\mathfrak{g}$ to $G$ as follows. Regard $\mathfrak{g}$ as the tangent space to the identity $e$ in $G$. For any $\xi \in \mathfrak{g}$ there exists a unique homomorphism $\gamma_\xi : \mathbb{R} \to G$ satisfying $\gamma_\xi(0) = e$ and $\dot{\gamma}_\xi(0) = \xi$ (see page 102 of [Warn]). We define

$$\exp(\xi) = \gamma_\xi(1).$$

The terminology and notation are motivated by the fact that, for the general linear group $GL(n, \mathbb{R})$,

$$\exp : gl(n, \mathbb{R}) \to GL(n, \mathbb{R})$$

is precisely matrix exponentiation (Example 3.35 of [Warn]). The same is true of any matrix group so in this case we will often use $\exp(\xi)$ and $e^\xi$ interchangeably.

\textit{Exercise 1.1.3.} Let $G$ be the additive group $\mathbb{R}^n$. Then the Lie algebra can be canonically identified with $\mathbb{R}^n$ (Exercise 1.1.1). Show that, with this identification, the exponential map is given by

$$\exp(\xi) = \xi.$$  

We will have occasion to need a more geometrical description of the rotation group $SO(3)$. The following is Lemma A.1 of [Nab2].

\textbf{Theorem 1.1.4.} Let $N$ be an element of $so(3)$. Then the matrix exponential $e^{tN}$ is in $SO(3)$ for every $t \in \mathbb{R}$. Conversely, if $A$ is any element of $SO(3)$, then there is a unique $t \in [0, \pi]$ and a unit vector $\mathbf{h} = (n^1, n^2, n^3)$ in $\mathbb{R}^3$ for which

$$A = e^{tN} = id_{3 \times 3} + (\sin t)N + (1 - \cos t)N^2,$$

where $N$ is the element of $so(3)$ given by
\[ N = \begin{pmatrix} 0 & -n^3 & n^2 \\ n^3 & 0 & -n^1 \\ -n^2 & n^1 & 0 \end{pmatrix}. \]

In particular, the exponential map on \( \text{so}(3) \) is surjective.

Geometrically, one thinks of \( A = e^{tN} \) as the rotation of \( \mathbb{R}^3 \) through \( t \) radians about an axis along \( \hat{n} \) in a sense determined by the right-hand rule from the direction of \( \hat{n} \).

If \( G \) is a group and \( M \) is a set, then a left action of \( G \) on \( M \) is a map \( \sigma : G \times M \to M \), generally written \( \sigma(g, x) = g \cdot x \), satisfying \( e \cdot x = x \) for every \( x \in M \), where \( e \in G \) is the identity element, and \( g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x \) for all \( g_1, g_2 \in G \) and all \( x \in M \). For each \( g \in G \) the map \( \sigma_g : M \to M \) defined by \( \sigma_g(x) = g \cdot x \) is then a bijection of \( M \) onto itself with inverse \( \sigma_{g^{-1}} \). The orbit \( O_{x_0} \) of \( x_0 \in M \) under this action is defined by

\[ O_{x_0} = \{ g \cdot x_0 : g \in G \} \]

and the isotropy subgroup \( H_{x_0} \) if \( x_0 \) is

\[ H_{x_0} = \{ g \in G : g \cdot x_0 = x_0 \}. \]

The action is said to be transitive if \( O_{x_0} = M \) for every \( x_0 \in M \) and free if \( H_{x_0} = \{ e \} \) for every \( x_0 \in M \). If \( G \) is a topological group and \( M \) is a topological space, then the action \( \sigma : G \times M \to M \) is required to be continuous and it follows that each \( \sigma_g \) is a homeomorphism. If \( G \) is a Lie group and \( M \) is a smooth manifold, then \( \sigma : G \times M \to M \) is required to be smooth and it follows that each \( \sigma_g \) is a diffeomorphism.

Similarly, a right action of \( G \) on \( M \) is a map \( \tau : M \times G \to M \), denoted \( \tau(x, g) = x \cdot g \), that satisfies \( x \cdot e = x \) and \( (x \cdot g_1) \cdot g_2 = x \cdot (g_1 g_2) \) for all \( x \in M \) and all \( g_1, g_2 \in G \). Given a right action \( \tau \) one can define a left action \( \sigma \) by \( \sigma(g, x) = \tau(x, g^{-1}) \). Similarly, every left action gives rise to a right action. All of the definitions we have introduced for left actions have obvious analogues for right actions. We will see an abundance of examples of such group actions as we proceed so for the moment we will content ourselves with just the following.

**Example 1.1.5.** Let \( G \) be a matrix Lie group and identify its Lie algebra \( \mathfrak{g} \) with the tangent space \( T_{id}(G) \) at the identity. \( G \) acts on itself by conjugation, that is, the map \( \sigma : G \times G \to G \) defined by \( \sigma(g, h) = g \cdot h = ghg^{-1} \) is a smooth left action of \( G \) on \( G \).

**Remark 1.1.4.** The corresponding right action \( \tau : G \times G \to G \) is given by \( \tau(h, g) = h \cdot g = g^{-1}hg \). Everything that follows has an obvious analogue for this right action which we will leave it to you to write out.

For each fixed \( g \in G \), the map \( \sigma_g : G \to G \) defined by

\[ \sigma_g(h) = ghg^{-1} \]
1.1 Lie Groups

for all $h \in G$ is a diffeomorphism. Its derivative at the identity is denoted $Ad_g$.

$$Ad_g = (\sigma_g)_{id} : g \to g$$

We claim that, for every $X \in \mathfrak{g}$,

$$Ad_g(X) = gXg^{-1}.$$ 

Indeed, since $t \mapsto e^{tX}$ is a smooth curve in $G$ through $id$ with velocity vector $X$ at $t = 0$,

$$Ad_g(X) = (\sigma_g)_{id}(X) = \frac{d}{dt}(ge^{tX}g^{-1})|_{t=0} = g\left(\frac{d}{dt}e^{tX}|_{t=0}\right)g^{-1} = gXg^{-1}.$$

Thus, $G$ acts on its Lie algebra $\mathfrak{g}$ by conjugation.

**Exercise 1.1.4.** Show that, for any $g \in G$ and any $X, Y \in \mathfrak{g}$,

$$Ad_g([X,Y]) = [Ad_g(X), Ad_g(Y)].$$

Each $Ad_g$ is therefore a Lie algebra homomorphism from $\mathfrak{g}$ to $\mathfrak{g}$. Since $Ad_g$ is clearly bijective, it is a Lie algebra isomorphism of $\mathfrak{g}$ onto itself, that is, an automorphism of $\mathfrak{g}$. The set $Aut(\mathfrak{g})$ of all automorphisms of $\mathfrak{g}$ is a closed subgroup of the general linear group $GL(\mathfrak{g})$ of $\mathfrak{g}$ and is therefore a Lie group. The map

$$Ad : G \to Aut(\mathfrak{g})$$

that sends $g \in G$ to $Ad_g \in Aut(\mathfrak{g})$ is smooth and its derivative at the identity is a linear map from $\mathfrak{g}$ to the Lie algebra of $Aut(\mathfrak{g})$. This map is denoted

$$ad = Ad_{id}$$

and its value at any $X \in \mathfrak{g}$ is denoted $ad_X = Ad_{id}(X)$. It follows from the Jacobi identity that each $ad_X$ is a **derivation** of $\mathfrak{g}$, that is, a linear map that satisfies the **Leibniz rule**

$$ad_X[Y,Z] = [Y,ad_XZ] + [ad_XY,Z]$$

for all $Y, Z \in \mathfrak{g}$. Thus,

$$ad : \mathfrak{g} \to Der(\mathfrak{g}) \subseteq End(\mathfrak{g}),$$

where $Der(\mathfrak{g})$ is the linear space of all derivations of $\mathfrak{g}$ (which is, in fact, the Lie algebra of $Aut(\mathfrak{g})$). We claim that, for any $Y \in \mathfrak{g}$, the value of $ad_X$ at $Y$ is given by

$$ad_XY = [X,Y].$$
Indeed, since $t \mapsto e^{tX}$ is a smooth curve in $G$ through $id$ with velocity vector $X$ at $t = 0$,

$$ad_X Y = [Ad_{id}(X)](Y) = \left. \frac{d}{dt} (Ad_{e^{tX}}(Y)) \right|_{t=0} = \left. \frac{d}{dt} (e^{tX} Y e^{-tX}) \right|_{t=0}$$

$$= e^{tX} \left. \frac{d}{dt} (Ye^{-tX}) \right|_{t=0} + \left. \frac{d}{dt} (e^{tX}) \right|_{t=0} (Ye^{-tX}) \right|_{t=0}$$

$$= -YX + XY$$

$$= [X,Y].$$

The map $ad$ is called the adjoint representation of $\mathfrak{g}$ and $ad_X Y = [X,Y]$ is the adjoint action of $X$ on $Y$. The Lie algebra homomorphism $ad : \mathfrak{g} \to \text{End}(\mathfrak{g})$ is a Lie algebra representation of $\mathfrak{g}$ on $\mathfrak{g}$.

Now let $G$ be a Lie group, $H$ a closed (not necessarily normal) subgroup of $G$, and $\pi : G \to G/H$ the natural projection of $G$ onto the set of left cosets of $H$ in $G$.

$$\pi(g_0) = [g_0] = g_0 H$$

Supply $G/H$ with the quotient topology determined by $\pi$. Then it is not difficult to show that $\pi$ is a continuous, open map and the left action of $G$ on $G/H$ defined by

$$(g, [g_0]) \in G \times G/H \mapsto g \cdot [g_0] = [gg_0] \in G/H$$

is continuous. The action is also transitive on $G/H$, that is, for any two points $[g_0]$ and $[g_1]$ in $G/H$ there is a $g \in G$ for which $[g_1] = g \cdot [g_0]$, namely, $g = g_1 g_0^{-1}$. For this reason $G/H$ is called a homogeneous space. The corresponding statements in the differentiable category require considerably more work and are contained in the following result which combines Theorem 3.58 and Theorem 3.64 of [Warn].

**Theorem 1.1.5.** Let $G$ be a Lie group, $H$ a closed subgroup of $G$ and $\pi : G \to G/H$ the natural projection of $G$ onto the space of left cosets of $H$ in $G$. Then $G/H$ admits a unique differentiable structure for which the left action $(g, [g_0]) \in G \times G/H \mapsto g \cdot [g_0] = [gg_0] \in G/H$ is smooth. Relative to this differentiable structure all of the following are satisfied.

1. $\pi : G \to G/H$ is smooth.
2. If $H$ is a normal subgroup of $G$, then $G/H$ is a Lie group.
3. $\pi : G \to G/H$ admits smooth local sections, that is, for each point $[g_0] \in G/H$ there exists an open neighborhood $U$ of $[g_0]$ in $G/H$ and a smooth map $s : U \to \pi^{-1}(U) \subseteq G$ such that $\pi \circ s = id_U$.

The third apart of Theorem 1.1.5 asserts that one can locally make a smooth selection of a representative from each equivalence class in $G/H$. We will put Theorem 1.1.5 to good use in Section 1.4 when we describe Mackey’s theory of induced representations. The same is true of the following, which is Theorem 3.62 of [Warn].
Theorem 1.1.6. Let $\sigma : G \times M \to M$, $(g, m) \mapsto g \cdot m$, be a smooth, transitive left action of the Lie group $G$ on the manifold $M$. Let $m_0$ be an arbitrary point of $M$ and $H_{m_0} = \{g \in G : g \cdot m_0 = m_0\}$ the isotropy subgroup of $m_0$ in $G$. Then $H_{m_0}$ is a closed subgroup of $G$ and the map

$$\beta_{m_0} : G/H_{m_0} \to M$$

$$\beta_{m_0}([g]) = g \cdot m_0$$

is a diffeomorphism of $G/H_{m_0}$ onto $M$.

Notice that $g \cdot m_0$ is independent of the representative $g$ chosen from $[g] \in G/H_{m_0}$ since $H_{m_0}$ is the isotropy group of $m_0$.

Example 1.1.6. The rotation group $\text{SO}(3)$ acts smoothly and transitively on the 2-sphere $S^2$ in the following way. Identify each matrix $A \in \text{SO}(3)$ with the matrix, relative to the standard basis for $\mathbb{R}^3$, of an orthogonal linear transformation of $\mathbb{R}^3$, which we will also denote $A$. Define $\sigma : \text{SO}(3) \times S^2 \to S^2$ by $\sigma(A, x) = A \cdot x = A(x)$. Then $\sigma$ is clearly a smooth left action of $\text{SO}(3)$ on $S^2$. This action is transitive on $S^2$ and the isotropy subgroup of the north pole in $S^2$ is isomorphic to $\text{SO}(2)$ (see pages 90-91 of [Nab2]). We conclude from Theorem 1.1.6 that $S^2$ is diffeomorphic to the homogeneous manifold $\text{SO}(3)/\text{SO}(2)$. All of this generalizes at once to prove

$$S^{n-1} \cong \text{SO}(n)/\text{SO}(n-1)$$

for any $n \geq 2$ ($\text{SO}(1)$ is the trivial group).

Finally, we will need (in Section 2.4) a result on covering spaces. Recall that if $M$ and $N$ are smooth manifolds and $F : M \to N$ is a smooth mapping of $M$ onto $N$, then $F$ is said to be a (smooth) covering map and $M$ is said to be a (smooth) covering space of $N$ if each point $x \in N$ has an open neighborhood $U$ in $N$ with the property that $F^{-1}(U)$ is a disjoint union of open sets in $M$ each of which is mapped diffeomorphically onto $U$ by $F$. If $M$ is simply connected, then it is called the universal covering space of $N$ because it covers any other covering space of $N$. The following is Proposition 9.30 of [Lee].

Theorem 1.1.7. Let $G$ and $H$ be connected Lie groups and $F : G \to H$ a smooth homomorphism. Then the following are equivalent.

1. $F$ is surjective and has discrete kernel.
2. $F$ is a smooth covering map.

In this case we refer to $G$ as a covering group of $H$. Since $G$ and $H$ are locally diffeomorphic, they have the same Lie algebras (see Exercise 1.2.6 (5) for a simple example). In Section 2.4 we will construct two important examples of universal covering groups.
1.2 Unitary Group Representations

Although our interest is almost exclusively in unitary representations of Lie groups on Hilbert spaces it will be convenient to begin in a more general context. We let $G$ denote an arbitrary group and $V$ a real or complex vector space. For the moment we impose no topological structure on either of these. Denote by $\text{GL}(V)$ the group, under composition, of all invertible linear transformations of $V$ onto itself. Then a homomorphism $\sigma : G \to \text{GL}(V)$ of $G$ into $\text{GL}(V)$ is called a representation of $G$ on $V$. The dimension of $V$ (whether finite or infinite) is called the dimension of the representation and the elements of $V$ are called carriers of the representation. Notice that a representation of $G$ on $V$ determines a left action of $G$ on $V$ defined by $g \cdot v = \sigma(g)(v)$ and that, conversely, any left action of $G$ on $V$ by linear transformations determines a representation of $G$ on $V$ by the same equation. A subspace $S$ of $V$ with the property that $S$ is invariant under every $\sigma(g)$ (meaning $\sigma(g)(S) \subseteq S \forall g \in G$), is said to be an invariant subspace for $\sigma$. If $G$ and $V$ are endowed with topologies, then some sort of continuity requirement is imposed on the representations of $G$ on $V$. We turn now to the case of most interest to us.

We let $G$ be a Lie group and denote its identity element $e_G$ or simply $e$ if no confusion will arise. Let $\mathcal{H}$ be a separable, complex Hilbert space (either finite- or infinite-dimensional) and $\mathcal{U}(\mathcal{H})$ the group of unitary operators on $\mathcal{H}$. A unitary representation of $G$ on $\mathcal{H}$ is a group homomorphism $\sigma : G \to \mathcal{U}(\mathcal{H})$.

$\sigma$ is strongly continuous if, for each fixed $v \in \mathcal{H}$, the map

$$g \to \sigma(g)v : G \to \mathcal{H}$$

is continuous in the norm topology of $\mathcal{H}$, that is,

$$g \to g_0 \text{ in } G \Rightarrow \|\sigma(g)v - \sigma(g_0)v\| \to 0 \text{ in } \mathbb{R}. \tag{1.1}$$

The representation is said to be trivial if it sends every $g \in G$ to the identity operator $id_{\mathcal{H}} = I$ on $\mathcal{H}$.

Exercise 1.2.1. Let $G$ be a matrix Lie group, $\mathcal{H}$ a separable, complex Hilbert space and $\sigma : G \to \mathcal{U}(\mathcal{H})$ a unitary representation. Show that the following are equivalent.

1. $\sigma$ is strongly continuous, that is, satisfies (1.1).
2. $\sigma$ is weakly continuous, that is, for all $u, v \in \mathcal{H},$

$$g \to g_0 \text{ in } G \Rightarrow \langle \sigma(g)u, v \rangle \to \langle \sigma(g_0)u, v \rangle \text{ in } \mathbb{C}. \tag{1.2}$$

3. For each $u \in \mathcal{H}$, the map $g \in G \to \langle \sigma(g)u, u \rangle \in \mathbb{C}$ is continuous at $e$.

Show also that if $\mathcal{H}$ is finite-dimensional, then all of these are equivalent to the continuity of $\sigma : G \to \mathcal{U}(\mathcal{H})$ when $\mathcal{U}(\mathcal{H})$ is given the operator norm topology.
Hint: For (3) \(\Rightarrow\) (1) show that 
\[ \|\sigma(g)u - \sigma(g_0)u\|^2 = 2\|u\|^2 - 2Re \langle\sigma(g_0^{-1})u, u\rangle \leq \|u\|^2 - \langle\sigma(g_0^{-1})u, u\rangle, \]

A linear subspace \(\mathcal{H}_0\) of \(\mathcal{H}\) is said to be invariant under \(G \to \mathcal{U}(\mathcal{H})\) if \(\sigma(g)(\mathcal{H}_0) \subseteq \mathcal{H}_0\) for every \(g \in G\). The zero subspace 0 and \(\mathcal{H}\) itself are always invariant. If \(\sigma : G \to \mathcal{U}(\mathcal{H})\) is nontrivial and if 0 and \(\mathcal{H}\) are the only closed invariant subspaces, then \(\sigma : G \to \mathcal{U}(\mathcal{H})\) is said to be irreducible. If there are closed invariant subspaces other than 0 and \(\mathcal{H}\), then the representation is said to be reducible. Two unitary representations \(\sigma_1 : G \to \mathcal{U}(\mathcal{H}_1)\) and \(\sigma_2 : G \to \mathcal{U}(\mathcal{H}_2)\) of \(G\) are said to be unitarily equivalent if there exists a unitary operator \(U : \mathcal{H}_1 \to \mathcal{H}_2\) of \(\mathcal{H}_1\) onto \(\mathcal{H}_2\) such that

\[ U\sigma_1(g) = \sigma_2(g)U \quad \forall g \in G, \]

that is,

\[ \sigma_2(g) = U\sigma_1(g)U^{-1} \quad \forall g \in G. \]

In this case, \(U\) is said to intertwine the representations \(\sigma_1\) and \(\sigma_2\).

Another item we will need is the following infinite-dimensional version of Schur’s Lemma. The proof is a nice application of the Spectral Theorem (see Section 5.5 of [Nab5]) and is not so readily available in the literature so we will provide the details.

Remark 1.2.1. A more general version of Schur’s Lemma is proved in Appendix 1 of [Lang4].

Theorem 1.2.1. (Schur’s Lemma) Let \(G\) be a Lie group, \(\mathcal{H}\) a separable, complex Hilbert space, and \(\sigma : G \to \mathcal{U}(\mathcal{H})\) a strongly continuous unitary representation of \(G\). Then \(\sigma : G \to \mathcal{U}(\mathcal{H})\) is irreducible if and only if the only bounded operators \(A : \mathcal{H} \to \mathcal{H}\) that commute with every \(\sigma(g)\)

\[ \sigma(g)A = A\sigma(g) \quad \forall g \in G \]

are those of the form \(A = cI\), where \(c\) is a complex number of modulus one and \(I\) is the identity operator on \(\mathcal{H}\).

Proof. Suppose first that the only bounded operators that commute with every \(\sigma(g)\) are constant multiples of the identity. We will show that the representation is irreducible. Let \(\mathcal{H}_0\) be a closed subspace of \(\mathcal{H}\) that is invariant under every \(\sigma(g)\).

Exercise 1.2.2. Show that the orthogonal complement \(\mathcal{H}_0^\perp\) of \(\mathcal{H}_0\) is also invariant under every \(\sigma(g)\).

Now, let \(P : \mathcal{H} \to \mathcal{H}_0\) be the orthogonal projection onto \(\mathcal{H}_0\).
Exercise 1.2.3. Show that $\sigma(g)P = P\sigma(g)$ for every $g \in G$.

According to our assumption, $P$ is a constant multiple of the identity. Being a projection, $P^2 = P$ so the constant is either 0 or 1. Thus, $\mathcal{H}_0 = P(\mathcal{H})$ is either 0 or $\mathcal{H}$, as required.

Now, for the converse we will assume that $\sigma : G \to \mathcal{U}(\mathcal{H})$ is irreducible and that $A : \mathcal{H} \to \mathcal{H}$ is a bounded operator that commutes with every $\sigma(g)$. We must show that $A$ is a constant multiple of the identity operator on $\mathcal{H}$. Let $A^* : \mathcal{H} \to \mathcal{H}$ denote the adjoint of $A$ (also a bounded operator on $\mathcal{H}$).

Exercise 1.2.4. Show that $A^*$ also commutes with every $\sigma(g)$.

Notice that $\frac{1}{2}(A + A^*)$ and $\frac{i}{2}(A - A^*)$ are both self-adjoint and both commute with every $\sigma(g)$. Moreover,

$$A = \frac{1}{2}(A + A^*) + \frac{1}{i} \frac{1}{2}(A - A^*)$$

Consequently, it will be enough to prove that bounded self-adjoint operators that commute with every $\sigma(g)$ must be constant multiples of the identity. Accordingly, we may assume that $A$ is self-adjoint. Then, by the Spectral Theorem, $A$ has associated with it a unique spectral measure $E^A$. Moreover, since $A$ commutes with every $\sigma(g)$, so does $E^A(S)$ for any Borel set $S \subseteq \mathbb{R}$. From this it follows that each closed linear subspace $E^A(S)(\mathcal{H})$ is invariant under $\sigma : G \to \mathcal{U}(\mathcal{H})$. But, by irreducibility, this means that

$$E^A(S)(\mathcal{H}) = 0 \text{ or } E^A(S)(\mathcal{H}) = \mathcal{H}$$

for every Borel set $S$ in $\mathbb{R}$.

Since $A$ is bounded there exist $a_1 < b_1$ in $\mathbb{R}$ such that, if $S \cap [a_1, b_1] = \emptyset$, then $E^A(S) = 0$. In particular, $E^A([a_1, b_1]) = I$. Write

$$[a_1, b_1] = \left[ a_1, \frac{a_1 + b_1}{2} \right] \cup \left[ \frac{a_1 + b_1}{2}, b_1 \right].$$

Now notice that, if $E^A((\frac{a_1 + b_1}{2})) = I$, then the Spectral Theorem gives $A = \frac{a_1 + b_1}{2} I$ and we are done. Otherwise, $E^A$ must be $I$ on one of the intervals and 0 on the other. Denote by $[a_2, b_2]$ the interval on which it is $I$. Applying the same argument to $[a_2, b_2]$ we either prove the result (at the midpoint) or we obtain an interval $[a_3, b_3]$ of half the length of $[a_2, b_2]$ on which $E^A$ is $I$. Continuing inductively, we either prove the result in a finite number of steps or we obtain a nested sequence $[a_1, b_1] \supseteq [a_2, b_2] \supseteq [a_3, b_3] \supseteq \cdots$ of intervals whose lengths approach zero and for which $E^A([a_i, b_i]) = I$ for every $i = 1, 2, 3, \ldots$. By the Cantor Intersection Theorem (Theorem C, page 73, of [Simm1]), $\bigcap_{i=1}^{\infty} [a_i, b_i] = \{ c \}$ for some $c \in \mathbb{R}$. Since $E^A(\mathbb{R} - [a_i, b_i]) = 0$ for each $i$, $0 = E^A(\bigcup_{i=1}^{\infty}(\mathbb{R} - [a_i, b_i])) = E^A(\mathbb{R} - \bigcap_{i=1}^{\infty}[a_i, b_i]) = E^A(\mathbb{R} - \{ c \})$. Thus, $E^A(\{ c \}) = I$ so again we have $A = cI$. \(\Box\)
Corollary 1.2.2. Let $A$ be an Abelian Lie group. Then every irreducible, unitary representation of $A$ on a complex, separable Hilbert space $\mathcal{H}$ is 1-dimensional.

Exercise 1.2.5. Prove Corollary 1.2.2.

It is the unitary representations of the Poincaré group and its universal cover (Section 2.4) that are of most interest to us, but to describe these we will need not only all of the machinery of the next three sections, but also an explicit description of the irreducible, unitary representations of SU(2). We will conclude this section with the latter.

Example 1.2.1. The special unitary group SU(2) consists of all $2 \times 2$ complex matrices $U$ that are unitary ($UU^T = U^T U = \text{id}_{2\times2}$) and have determinant $\det(U) = 1$. Every $U \in \text{SU}(2)$ can be written uniquely in the form

$$U = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix},$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$ (Lemma 1.1.3 of [Nab2]). The inverse of $U$ is given by

$$U^{-1} = \begin{pmatrix} \bar{\alpha} & -\bar{\beta} \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}.$$ 

SU(2) is a group under matrix multiplication. Indeed, it is a closed subgroup of the Lie group GL($2, \mathbb{C}$) of invertible $2 \times 2$ complex matrices and is therefore a matrix Lie group. The map

$$(\alpha, \beta) \mapsto \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \in \text{SU}(2)$$

identifies SU(2) topologically with the 3-sphere $S^3$ (see Section 1.1 of [Nab2]). In particular, SU(2) is compact (being closed and bounded in $\mathbb{R}^4$) and simply connected (see pages 118-119 of [Nab2]). It follows from compactness that every continuous, irreducible representation of SU(2) is finite-dimensional (see Theorem 5.2 of [Fol2]). Another consequence of compactness is that every continuous, finite-dimensional representation $\sigma : SU(2) \to GL(V)$ of SU(2) on a complex vector space $V$ is “unitarizable” in the sense that there exists a Hermitian inner product on $V$ with respect to which $\sigma$ is unitary (see page 128 of [Fol2]). As a result it will be enough to consider the continuous, finite-dimensional, irreducible, unitary representations of SU(2). It is our good fortune that all of these are known.

We will begin by just writing out a few nontrivial, finite-dimensional representations of SU(2). The most obvious of these is the standard, or defining representation $\tau$ of SU(2) on $\mathbb{C}^2$ that simply identifies any $U \in \text{SU}(2)$ with a linear transformation on $\mathbb{C}^2$ by matrix multiplication, that is,
\[ \tau(U) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = U \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \alpha z_1 + \beta z_2 \\ -\beta z_1 + \alpha z_2 \end{pmatrix}. \]

**Remark 1.2.2.** We will feel free to regard the elements of \( \mathbb{C}^n \) as either \( n \)-tuples or column vectors of complex numbers and will write them as \( Z = (z_1, \ldots, z_n) \) or \( Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \). In any case, the standard Hermitian inner product \( \langle Z, W \rangle \) of \( Z \) and \( W \) in \( \mathbb{C}^n \) is \( \sum z_i w_i \). Naturally, this turns \( \mathbb{C}^n \) into a complex Hilbert space. With the topology determined by the corresponding norm, \( \mathbb{C}^n \) is homeomorphic to \( \mathbb{R}^{2n} \) and \( \tau \) is continuous.

There is another obvious representation of \( \text{SU}(2) \) on \( \mathbb{C}^2 \) called the *conjugation representation* which matrix multiplies by \( U \) rather than \( \bar{U} \). We will denote this \( \bar{\tau} \) so that

\[ \bar{\tau}(U) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \bar{U} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \bar{\alpha} & \bar{\beta} \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \bar{\alpha} z_1 + \bar{\beta} z_2 \\ -\bar{\beta} z_1 + \bar{\alpha} z_2 \end{pmatrix}. \]

We would like to show, however, that \( \bar{\tau} \) is not really anything new since it is unitarily equivalent to \( \tau \). To prove this we will use some properties of the so-called “Pauli spin matrices” which we introduce in the following Exercise. We will try to give some sense of where they come from and why they are interesting.

**Exercise 1.2.6.** Denote by \( \mathcal{R}^3 \) the set of all \( 2 \times 2 \) complex, Hermitian matrices \( X \) \((X^T = X)\) with trace zero \((\text{Trace}(X) = 0)\).

1. Show that every \( X \in \mathcal{R}^3 \) can be uniquely written as

\[ X = \begin{pmatrix} x^3 & x^1 - i x^2 \\ x^1 + i x^2 & -x^3 \end{pmatrix} = x^1 \sigma_1 + x^2 \sigma_2 + x^3 \sigma_3, \]

where \( x^1, x^2 \) and \( x^3 \) are real numbers and

\[ \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \]

are the so-called *Pauli spin matrices*. Show that these are all unitary and equal to their own inverses.

2. Show that, with the operations of matrix addition and (real) scalar multiplication, \( \mathcal{R}^3 \) is a 3-dimensional, real vector space and \( \{ \sigma_1, \sigma_2, \sigma_3 \} \) is a basis. Consequently, \( \mathcal{R}^3 \) is linearly isomorphic to \( \mathbb{R}^3 \). Furthermore, defining an orientation on \( \mathcal{R}^3 \) by decreeing that \( \{ \sigma_1, \sigma_2, \sigma_3 \} \) is an oriented basis, the map \( X \to (x^1, x^2, x^3) \) is an orientation preserving isomorphism when \( \mathcal{R}^3 \) is given its usual orientation.
3. Show that
\[ \sigma_1 \sigma_2 = i \sigma_3, \quad \sigma_2 \sigma_3 = i \sigma_1, \quad \sigma_3 \sigma_1 = i \sigma_2, \quad \sigma_1 \sigma_2 \sigma_3 = i I, \]
where \( I \) is the 2 \times 2 identity matrix.

4. Show that \( \sigma_1, \sigma_2, \sigma_3 \) satisfy the \textit{commutation relations}
\[ [\sigma_1, \sigma_2]_-= 2 i \sigma_3, \quad [\sigma_2, \sigma_3]_- = 2 i \sigma_1, \quad [\sigma_3, \sigma_1]_- = 2 i \sigma_2, \]
where \([, ]_-\) denotes the matrix commutator \(([A, B]_- = AB - BA)\).

5. Show that \( \sigma_1, \sigma_2, \sigma_3 \) satisfy the \textit{anticommutation relations},
\[ [\sigma_i, \sigma_j]_+ = 2 \delta_{ij} I, \quad i, j = 1, 2, 3, \]
where \([, ]_+\) denotes the matrix anticommutator \(([A, B]_+ = AB + BA)\) and \(\delta_{ij}\) is the Kronecker delta. In particular, each \(\sigma_i\) squares to the identity matrix.

6. Show that, if \(X = x^1 \sigma_1 + x^2 \sigma_2 + x^3 \sigma_3\) and \(Y = y^1 \sigma_1 + y^2 \sigma_2 + y^3 \sigma_3\), then
\[ \frac{1}{2} [X, Y]_+ = (x^1 y^1 + x^2 y^2 + x^3 y^3) I. \]

Conclude that, if one defines an inner product \(\langle X, Y \rangle_{\mathbb{R}^3}\) on \(\mathbb{R}^3\) by
\[ \frac{1}{2} [X, Y]_+ = \langle X, Y \rangle_{\mathbb{R}^3} I, \]
then \(\{\sigma_1, \sigma_2, \sigma_3\}\) is an \textit{oriented, orthonormal basis for} \(\mathbb{R}^3\) and \(\mathbb{R}^3\) is isometric to \(\mathbb{R}^3\). We will refer to \(\mathbb{R}^3\) as the \textit{spin model} of \(\mathbb{R}^3\).

7. Regard the matrices \(\sigma_1, \sigma_2, \sigma_3\) as linear operators on \(\mathbb{C}^2\) (as a 2-dimensional, complex vector space with its standard Hermitian inner product) and show that each of these operators has eigenvalues \(\pm 1\) with normalized, orthogonal eigenvectors given as follows.

\[ \sigma_1 : \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \]
\[ \sigma_2 : \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \]
\[ \sigma_3 : \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \]

8. Show that, for any \(U \in SU(2)\),
\[ \overline{U} = \sigma_3 U \sigma_2^{-1} \]
and conclude that the conjugation representation \(\tau\) and the standard representation \(\tau\) are unitarily equivalent.
9. Show that \( u_j = -\frac{1}{2} \sigma_j, j = 1, 2, 3 \), is a basis for the Lie algebra \( \mathfrak{su}(2) \) of \( SU(2) \) relative to which the structure constants are given by

\[
[u_j, u_k] = \epsilon_{jk} u_l, \quad j, k = 1, 2, 3
\]

and that the linear map from \( \mathfrak{so}(3) \) to \( \mathfrak{su}(2) \) defined by \( X_j \mapsto u_j, j = 1, 2, 3 \), is an isomorphism of Lie algebras (see Example 1.1.2).

10. Define the standard representation \( \tau \) and the conjugation representation \( \overline{\tau} \) of \( SL(2, \mathbb{C}) \) in exactly the same way as for \( SU(2) \) and show that these are not unitarily equivalent. Hint: Let

\[
U_0 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.
\]

Then \( U_0 \in SU(2) \subseteq SL(2, \mathbb{C}) \) so \( \sigma_2 U_0 \sigma_2^{-1} = \overline{U}_0 \). Show that \( \sigma_2 \) and its nonzero, real scalar multiples are the only complex, \( 2 \times 2 \) matrices for which this is true. Now find an element \( A \) of \( SL(2, \mathbb{C}) \) for which \( \sigma_2 A \sigma_2^{-1} \neq \overline{A} \).

Now let’s make a general observation. Suppose we have a representation \( \sigma : G \to GL(\mathbb{C}^V) \) of \( G \) on some finite-dimensional complex vector space \( V \). Let \( \mathcal{F}(\mathbb{C}^V) \) be the vector space of all complex-valued functions \( f : \mathbb{C}^V \to \mathbb{C} \) on \( V \). Now define \( \lambda : G \to GL(\mathcal{F}(\mathbb{C}^V)) \) by

\[
(\lambda(g)f)(v) = f(\sigma(g)^{-1}(v))
\]

for all \( g \in G \) and all \( v \in \mathbb{C}^V \). Then \( \lambda \) is also a representation of \( G \) because

\[
\begin{align*}
(\lambda(g_1 g_2)f)(v) &= f(\sigma(g_1 g_2)^{-1}(v)) \\
&= f(\sigma(g_2)^{-1}\sigma(g_1)^{-1}(v)) \\
&= (\lambda(g_2)f)(\lambda(g_1)f)(v)
\end{align*}
\]

and so \( \lambda(g_1 g_2) = \lambda(g_1) \lambda(g_2) \). The same is true if \( \mathcal{F}(\mathbb{C}^V) \) is replaced by any vector space of complex-valued functions on \( \mathbb{C}^V \) with the property that it contains \( f(\sigma(g)^{-1}(v)) \) whenever it contains \( f(v) \). This suggests a means of producing new representations from given representations that we will now apply to \( SU(2) \).

We begin with the standard representation \( \tau : SU(2) \to GL(\mathbb{C}^2) \) of \( SU(2) \) on \( \mathbb{C}^2 \). For any integer \( j \geq 0 \) we let \( \mathcal{V}_j \) denote the complex vector space of all homogeneous polynomial functions of degree \( j \) in the two complex variables \( z_1 \) and \( z_2 \) (\( \mathcal{V}_0 \) consists of the constant functions of \( z_1 \) and \( z_2 \) so one can identify it with \( \mathbb{C} \)). Thus, any element of \( \mathcal{V}_j \) is of the form

\[
f(z_1, z_2) = a_0 z_1^j + a_1 z_1^{j-1} z_2 + a_2 z_1^{j-2} z_2^2 + \cdots + a_j z_2^j.
\]

The complex dimension of \( \mathcal{V}_j \) is \( j + 1 \). Define \( \tau_j : SU(2) \to GL(\mathcal{V}_j) \) by
We will describe this in the following example.

Generally we will find it more convenient to simply identify

\[ \mathcal{V}_j = \mathbb{C}^{j+1} \]

via the isomorphism described above and regard \( \tau_j \) as acting on the coefficients \((a_0, a_1, a_2, \ldots, a_j)\). Doing this in the notation we have employed up to this point is rather cumbersome so we will introduce a new way of writing all of this that is more convenient and also more in line with what one is likely to see in the physics literature. We will describe this in the following example.

**Example 1.2.2.** A typical element of \( \mathcal{V}_1 \) is a homogeneous polynomial

\[ f(z_1, z_2) = a_0 z_1 + a_1 z_2 \]

of degree 1 with complex coefficients. We now prefer to write this in the form

\[ f(z_1, z_2) = \sum_{A=1}^2 \xi^A z_A, \]

where \( \xi^A \) are complex numbers.
where $\xi^1 = a_0$ and $\xi^2 = a_1$. With the summation convention this is just

$$\xi^A z_A.$$ 

We are interested in how the coefficients $\xi^A$ “transform” $\xi^A \rightarrow \hat{\xi}^A$ under the action of $\tau_1$. For this we will write the entries of $U \in SU(2)$ as

$$U = \begin{pmatrix} U^1_1 & U^2_1 \\ U^2_1 & U^2_2 \end{pmatrix}$$

so that

$$U^{-1} = U^T = \begin{pmatrix} U^1_1 & U^2_1 \\ U^2_1 & U^2_2 \end{pmatrix}.$$

Now we compute

$$(\tau_1(U)f)(z_1, z_2) = f(\tau(U)^{-1}(z_1, z_2)) = f(U^1_1 z_1 + U^2_1 z_2, U^1_2 z_1 + U^2_2 z_2)$$

$$= \xi^1 (U^1_1 z_1 + U^2_1 z_2) + \xi^2 (U^1_2 z_1 + U^2_2 z_2)$$

$$= (U^1_1 \xi^1 + U^1_2 \xi^2) z_1 + (U^2_1 \xi^1 + U^2_2 \xi^2) z_2$$

$$= \hat{\xi}^A z_A,$$

where the transformed coordinates $\hat{\xi}^A$ are given by

$$\begin{pmatrix} \hat{\xi}^1 \\ \hat{\xi}^2 \end{pmatrix} = \begin{pmatrix} U^1_1 & U^2_1 \\ U^2_1 & U^2_2 \end{pmatrix} \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix}.$$

This is just the conjugation representation of $SU(2)$ on $\mathbb{C}^2$. As we have seen in Exercise 1.2.6 (8), this is unitarily equivalent to the standard representation

$$\begin{pmatrix} \hat{\xi}^1 \\ \hat{\xi}^2 \end{pmatrix} = \begin{pmatrix} U^1_1 & U^1_2 \\ U^2_1 & U^2_2 \end{pmatrix} \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix}$$

of $SU(2)$ on $\mathbb{C}^2$ and it will simplify the notation if we deal with this equivalent representation instead. We would now like to write this as

$$\hat{\xi}^A = U^A_B \xi^B, \ A = 1, 2,$$

where $B$ is summed over $B = 1, 2$.

Now notice that a typical element of $V_2$ is a homogeneous polynomial $a_0 z_1^2 + a_1 z_1 z_2 + a_2 z_2^2$ of degree 2 with complex coefficients and that this can be written in the form

$$\sum_{A_1, A_2 = 1}^{2} \xi^{A_1 A_2} z_{A_1} z_{A_2},$$
where the coefficients \( \xi^{A_1 A_2} \in \mathbb{C} \) are symmetric in \( A_1 \) and \( A_2 \) (specifically, \( \xi^{11} = a_0, \xi^{22} = a_2, \xi^{12} = \xi^{21} = a_1/2 \)). With the summation convention this is just

\[
\xi^{A_1 A_2} z_{A_1} z_{A_2}.
\]

**Exercise 1.2.7.** Show that the action of \( \tau_2 \) on \( \mathcal{V}_2 \), when thought of as a transformation of the coefficients, is unitarily equivalent to the representation of \( \text{SU}(2) \) on the subspace of \( \mathbb{C}^4 \) consisting of all 4-tuples \( (\xi^{A_1 A_2})_{A_1, A_2 = 1, 2} \) that are symmetric in \( A_1 A_2 \) and given by

\[
\hat{\xi}^{A_1 A_2} = U^{A_1 B_1} U^{A_2 B_2} \xi^{B_1 B_2},
\]

where \( B_1, B_2 \) are summed over \( B_1, B_2 = 1, 2 \).

In general, the elements of \( \mathcal{V}_j \) can be written in the form

\[
\xi^{A_1 A_2 \cdots A_j} z_{A_1} z_{A_2} \cdots z_{A_j},
\]

where the coefficients \( \xi^{A_1 A_2 \cdots A_j} \in \mathbb{C}, A_1, A_2, \ldots, A_j = 1, 2 \), are symmetric under all permutations of \( A_1 A_2 \ldots A_j \). The action of \( \tau_j \) on \( \mathcal{V}_j \), when thought of as a transformation of the coefficients, is unitarily equivalent to the representation of \( \text{SU}(2) \) on the subspace of \( \mathbb{C}^{2^j} \) consisting of all \( 2^j \)-tuples \( (\xi^{A_1 A_2 \cdots A_j})_{A_1, A_2, \ldots, A_j = 1, 2} \) that are invariant under all permutations of \( A_1 A_2 \cdots A_j \) and given by

\[
\hat{\xi}^{A_1 A_2 \cdots A_j} = U^{A_1 B_1} U^{A_2 B_2} \cdots U^{A_j B_j} \xi^{B_1 B_2 \cdots B_j}, \quad A_1, A_2, \ldots, A_j = 1, 2,
\]

(1.3)

where \( B_1, B_2, \ldots, B_j \) are summed over \( B_1, B_2, \ldots, B_j = 1, 2 \).

**Exercise 1.2.8.** Show that when \( j = 2 \) the transformation law (1.3) can be written as

\[
\begin{pmatrix}
\hat{\xi}^{11} \\
\hat{\xi}^{12} \\
\hat{\xi}^{21} \\
\hat{\xi}^{22}
\end{pmatrix}
= \begin{pmatrix}
U^{11}_{11} & U^{11}_{12} \\
U^{12}_{11} & U^{12}_{12} \\
U^{21}_{11} & U^{21}_{12} \\
U^{22}_{11} & U^{22}_{12}
\end{pmatrix}
\otimes
\begin{pmatrix}
\xi^{11} \\
\xi^{12} \\
\xi^{21} \\
\xi^{22}
\end{pmatrix},
\]

where \( \hat{\xi}^{21} = \xi^{12} \) and \( \otimes \) means the matrix tensor product. Now generalize.

The transformation law (1.3) describes an irreducible, unitary representation of \( \text{SU}(2) \). It is called the **spinor representation** of \( \text{SU}(2) \) of weight \( j \) and is denoted \( \mathcal{D}^{(j/2)} \). The carriers of the representation are the \( 2^j \)-tuples \( (\xi^{A_1 A_2 \cdots A_j})_{A_1, A_2, \ldots, A_j = 1, 2} \) in \( \mathbb{C}^{2^j} \) that are invariant under all permutations of \( A_1 A_2 \cdots A_j \) and are called the components of a (contravariant) \( \text{SU}(2) \)-spinor of rank \( j \). The dimension of the space \( \mathcal{E}(\mathcal{D}^{(j/2)}) \) of carriers (that is, the dimension of the representation) is \( j + 1 \).

Every irreducible, unitary representation of \( \text{SU}(2) \) is unitarily equivalent to one and only one of the spinor representations \( \mathcal{D}^{(j/2)}, j = 0, 1, 2, \ldots \).
Remark 1.2.4. It is common to denote the spinor representations $D^{(j)}$, where $s = j/2$ is in $\{0, \frac{1}{2}, 1, \frac{3}{2}, \ldots \}$ and refer to $s$ as the spin of the representation.

Remark 1.2.5. One can think of an SU(2)-spinor intuitively in a way that is entirely analogous to the old fashioned way of thinking about a tensor. One has assigned to every orthonormal basis for $\mathbb{C}^2$ a set of $2^j$ complex components $\xi^{A_1\cdots A_j}$, $A_1, \ldots, A_j = 1, 2$, totally symmetric in $A_1 \ldots A_j$. If two such bases are related by $U = (U^A_B)_{A,B=1,2}$ in SU(2), then the corresponding components are related by the transformation law (1.3). It is not unreasonable to ask if such things are of any use to physics. Now, one is familiar with a great many physical quantities that are represented mathematically by tensors of various types (mass, velocity, momentum, elastic stress and so on) and these tensors are simply the carriers of various representations of SO(3). But notice that when $j$ is even, that is, when the representation $D^{(j/2)}$ of SU(2) has integral spin, $D^{(j/2)}(-U) = D^{(j/2)}(U)$ for every $U \in$ SU(2). We will see in Section 2.4 that SU(2) is the universal double cover of the rotation group SO(3) and will conclude from this that integral spin representations of SU(2) descend to representations of SO(3) and so their carriers can be identified with tensors. Half-integral spin representations of SU(2) do not descend to representations of SO(3) so their carriers are not tensors and represent something new. That there are, indeed, physical quantities that require such spinors for their mathematical description and that these quantities have unexpected and counterintuitive properties is the topic of Appendix B of [Nab4].

Remark 1.2.6. If $U$ is assumed to be only in SL(2, $\mathbb{C}$) rather than the subgroup SU(2), then (1.3) still determines an irreducible, unitary representation of SL(2, $\mathbb{C}$) and the carriers are called the components of a (contravariant, undotted) SL(2, $\mathbb{C}$)-spinor. In this case, however, these do not exhaust all of the irreducible, unitary representations of SL(2, $\mathbb{C}$). The representations of SL(2, $\mathbb{C}$) are discussed in much more detail in Chapter 3 and Appendix B of [Nab4].

1.3 Projective Representations

Let $\mathcal{H}$ be a complex, separable Hilbert space. For each $v \in \mathcal{H} - \{0\}$

$$[v] = C v = \{ \lambda v : \lambda \in \mathbb{C} \}$$

is the ray in $\mathcal{H}$ containing $v$. The projectivization of $\mathcal{H}$ is the set of all such rays and is denoted

$$\mathbb{P}(\mathcal{H}) = \{ [v] : v \in \mathcal{H} - \{0\} \}.$$ 

The map
\[ \pi_P : \mathcal{H} - \{0\} \to \mathbb{P}(\mathcal{H}) \]
\[ \pi_P(v) = [v] \]

is a surjection and we provide \( \mathbb{P}(\mathcal{H}) \) with the quotient topology it determines. Stated otherwise, \( \mathbb{P}(\mathcal{H}) \) is the quotient space of \( \mathcal{H} - \{0\} \) by the equivalence relation \( \sim \) defined by \( v_1 \sim v_2 \) if and only if \( v_2 = \lambda v_1 \) for some \( \lambda \in \mathbb{C} \). The topology on \( \mathbb{P}(\mathcal{H}) \) is Hausdorff and \( \pi_P \) is an open mapping. There are two other useful ways of viewing \( \mathbb{P}(\mathcal{H}) \).

1. Let \( S(\mathcal{H}) = \{ e \in \mathcal{H} : \|e\|_{\mathcal{H}} = 1 \} \) be the unit sphere in \( \mathcal{H} \) with its relative topology. Define an equivalence relation \( \sim \) on \( S(\mathcal{H}) \) by \( e_1 \sim e_2 \) if and only if \( e_2 = \lambda e_1 \) for some \( \lambda \in \mathbb{C} \) with \( |\lambda| = 1 \). Then \( \mathbb{P}(\mathcal{H}) \) is homeomorphic to the quotient space \( S(\mathcal{H})/\sim \) of \( S(\mathcal{H}) \) by \( \sim \). In particular, one can think of \( \mathbb{P}(\mathcal{H}) \) as the space of unit rays \( [e] = \{ \lambda e : \lambda \in \mathbb{C}, |\lambda| = 1 \} \) in \( \mathcal{H} \).

2. Let \( \text{End}(\mathcal{H}) \) be the Banach space of all bounded linear operators on \( \mathcal{H} \) with the operator norm. For each \( v \in \mathcal{H} - \{0\} \), let \( \pi_v \in \text{End}(\mathcal{H}) \) be the orthogonal projection of \( \mathcal{H} \) onto \([v]\). Then \( v \mapsto \pi_v \) is a continuous map of \( \mathcal{H} - \{0\} \) into \( \text{End}(\mathcal{H}) \) which depends only on \([v]\) and therefore induces a continuous map of \( \mathbb{P}(\mathcal{H}) \) into \( \text{End}(\mathcal{H}) \). It follows that the image of this map is given the relative topology from \( \text{End}(\mathcal{H}) \), then this is a homeomorphism of \( \mathbb{P}(\mathcal{H}) \) onto the image so we can identify

\[ \mathbb{P}(\mathcal{H}) = \{ \pi_v \in \text{End}(\mathcal{H}) : v \in \mathcal{H} - \{0\} \}. \]

At this point we know only that \( \mathbb{P}(\mathcal{H}) \) is a Hausdorff topological space, but it is, in fact, a topological Hilbert manifold and, if \( \mathcal{H} \) is infinite-dimensional, it is locally homeomorphic to \( \mathcal{H} \). One sees this in the following way. Let \( e \) be a unit vector in \( \mathcal{H} \) and \( e^\perp \) its orthogonal complement in \( \mathcal{H} \). Then \( e^\perp \) is a closed linear subspace of \( \mathcal{H} \) and consequently it is a Hilbert space in its own right so \( \mathbb{P}(e^\perp) \) is defined. The inclusion \( e^\perp \hookrightarrow \mathcal{H} \) induces a continuous embedding of \( \mathbb{P}(e^\perp) \) into \( \mathbb{P}(\mathcal{H}) \) and we will identify \( \mathbb{P}(e^\perp) \) with its image. Thus, \( U_e = \mathbb{P}(\mathcal{H}) - \mathbb{P}(e^\perp) \) is an open neighborhood of \( \pi_P(e) \) in \( \mathbb{P}(\mathcal{H}) \) which we claim is homeomorphic to \( e^\perp \) (and this is isometrically isomorphic to \( \mathcal{H} \) if \( \mathcal{H} \) is infinite-dimensional). Each element of \( U_e \) is a ray \([v]\) for which \( \langle e, v \rangle_{\mathcal{H}} \neq 0 \) and we can define \( \chi([v]) \in e^\perp \) by

\[ \chi([v]) = \frac{v - \langle e, v \rangle_{\mathcal{H}} e}{\langle e, v \rangle_{\mathcal{H}}}. \]

Note that \( \chi([v]) \) clearly depends only on \([v]\) and is, indeed, in \( e^\perp \).

**Exercise 1.3.1.** Show that \( \chi \) is bijective with inverse given by

\[ \chi^{-1}(w) = [e + w] \]

for every \( w \in e^\perp \).

Both \( \chi \) and \( \chi^{-1} \) are continuous so \( \chi \) is a homeomorphism. Since the unit vector \( e \) was arbitrary the result follows.
Next we need to introduce the automorphism group \( \text{Aut}(\mathcal{H}) \) of \( \mathcal{H} \). Motivated by the physical interpretation in quantum mechanics (see page 235 of [Nab5]) we define, for any \( v, w \in \mathcal{H} - \{0\} \), the transition probability \(([v], [w])\) from state \([v]\) to state \([w]\) by

\[
([v], [w]) = (|[v], [w]|)_{\mathcal{H}} = \frac{|\langle v, w \rangle_{\mathcal{H}}|^2}{|v|_{\mathcal{H}}^2}.
\]

Notice that the right-hand side is independent of the representatives \( v \) and \( w \) chosen for the rays. Let \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) be two complex, separable Hilbert spaces. A map \( T : \mathbb{P}(\mathcal{H}_1) \to \mathbb{P}(\mathcal{H}_2) \) is an isomorphism of their projectivizations if it is a homeomorphism and preserves transition probabilities in the sense that \((|[v], [w]|)_{\mathcal{H}_1} = (T([v]), T([w]))_{\mathcal{H}_2}\) for all \( v, w \in \mathcal{H}_1 - \{0\} \). An automorphism of \( \mathbb{P}(\mathcal{H}) \) is an isomorphism of \( \mathbb{P}(\mathcal{H}) \) onto itself. The set of all such is denoted \( \text{Aut}(\mathbb{P}(\mathcal{H})) \) and is a group under composition. We will supply \( \text{Aut}(\mathbb{P}(\mathcal{H})) \) with a topology shortly, but first we need to describe an important result due to Wigner [Wig1]. For this we recall that a map \( U : \mathcal{H}_1 \to \mathcal{H}_2 \) is anti-unitary if it is additive \((U(v + w) = Uv + Uw \text{ for all } v, w \in \mathcal{H}_1)\), anti-linear \((U(Av) = \bar{A}U(v) \text{ for all } A \in \mathbb{C} \text{ and all } v \in \mathcal{H}_1)\), and satisfies \(\langle Uv, Uw \rangle_2 = \langle v, w \rangle_1\) for all \( v, w \in \mathcal{H}_1 \). A map \( U \) that is either unitary or anti-unitary induces an isomorphism \( T_U \) on the projectivizations via \( T_U([v]) = [U(v)] \). Wigner’s Theorem asserts that every isomorphism \( T : \mathbb{P}(\mathcal{H}_1) \to \mathbb{P}(\mathcal{H}_2) \) arises in this way from a map from \( \mathcal{H}_1 \) to \( \mathcal{H}_2 \) that is either unitary or anti-unitary. In fact, the result is more general in that it does not require the topological assumptions. For a proof of the following result one can consult [Barg]; [VDB] contains a shorter proof, but one that assumes continuity.

**Remark 1.3.1.** Wigner defined a symmetry of the quantum system whose Hilbert space is \( \mathcal{H} \) to be a bijection \( T : \mathbb{P}(\mathcal{H}) \to \mathbb{P}(\mathcal{H}) \) that preserves transition probabilities. Note that no continuity assumption is made. The following result implies, in particular, that every symmetry of a quantum system arises from an operator on \( \mathcal{H} \) that is either unitary or anti-unitary. One gets continuity (and much more) for free. We will return to the discussion of symmetries and their physical significance in Section 2.7.

**Theorem 1.3.1.** (Wigner’s Theorem on Symmetries) Let \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) be complex, separable Hilbert spaces and \( T : \mathbb{P}(\mathcal{H}_1) \to \mathbb{P}(\mathcal{H}_2) \) a mapping of \( \mathbb{P}(\mathcal{H}_1) \) into \( \mathbb{P}(\mathcal{H}_2) \) satisfying

\[
(T([v]), T([w]))_{\mathcal{H}_2} = ([v], [w])_{\mathcal{H}_1}
\]

for all \([v], [w] \in \mathbb{P}(\mathcal{H}_1)\). Then there exists a mapping \( U : \mathcal{H}_1 \to \mathcal{H}_2 \) satisfying

1. \( U(v) \in T([v]) \) for all \( v \in \mathcal{H}_1 - \{0\} \),

2. \( U(v + w) = U(v) + U(w) \) for all \( v, w \in \mathcal{H}_1 \), and
3. either
   a. \( U(\lambda v) = \lambda U(v) \) and \( \langle U(v), U(w) \rangle_{\mathcal{H}_1} = \langle v, w \rangle_{\mathcal{H}_1} \), or
   b. \( U(\lambda v) = \overline{\lambda} U(v) \) and \( \langle U(v), U(w) \rangle_{\mathcal{H}_1} = \overline{\langle v, w \rangle}_{\mathcal{H}_1} \)

Furthermore, if \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) are of dimension at least 2 and \( T([e_1]) = T([e_2]) \), where \( e_1 \) and \( e_2 \) are unit vectors in \( \mathcal{H}_1 \), then there is a unique such mapping \( U : \mathcal{H}_1 \to \mathcal{H}_2 \) for which \( U(e_1) = U(e_2) \).

In particular, every automorphism \( T \) of \( \mathbb{P}(\mathcal{H}) \) arises from an operator \( U \) on \( \mathcal{H} \) that is either unitary or anti-unitary via \( T([v]) = [Uv] \).

We denote by \( \mathcal{U}(\mathcal{H}) \) the group of unitary operators on \( \mathcal{H} \) with the strong operator topology. The set \( \mathcal{U}^-(\mathcal{H}) \) of anti-unitary operators, also with the strong operator topology, is not a group, but the product (composition) of two anti-unitary operators is unitary so the disjoint union \( \mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^-(\mathcal{H}) \) is a group under composition. With the strong operator topology, \( \mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^-(\mathcal{H}) \) is a Hausdorff topological group and \( \mathcal{U}(\mathcal{H}) \) is a closed and open subgroup (the component containing the identity operator \( I \)). We identify the group \( S^1 \) of complex numbers of modulus one with a subgroup of \( \mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^-(\mathcal{H}) \) via the map \( z \mapsto zI \) for \( z \in S^1 \). The following result gives an explicit description of \( \text{Aut}(\mathbb{P}(\mathcal{H})) \) and is the tool we need to provide \( \text{Aut}(\mathbb{P}(\mathcal{H})) \) with a natural topology.

**Proposition 1.3.2.** Let \( \mathcal{H} \) be a complex, separable Hilbert space of dimension at least two. Then the map

\[
Q : \mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^-(\mathcal{H}) \to \text{Aut}(\mathbb{P}(\mathcal{H}))
\]

defined by

\[
Q(U)([v]) = U([v]) = C(Uv)
\]

is a surjective group homomorphism with kernel \( S^1 \). Consequently, \( \text{Aut}(\mathbb{P}(\mathcal{H})) \) is isomorphic as a group to the quotient \( [\mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^-(\mathcal{H})]/S^1 \).

**Proof.** \( Q \) is a homomorphism because

\[
Q(U_1U_2)([v]) = [U_1(U_2v)] = [U_1(U_2v)] = Q(U_1)([U_2v]) = Q(U_1)Q(U_2)([v]) = Q(U_1)Q(U_2)([v]).
\]

\( Q \) is surjective by Wigner’s Theorem 1.3.1. The kernel of \( Q \) contains \( S^1 \) because, for every \( z \in S^1 \),

\[
Q(zI)([v]) = [zv] = [v].
\]

Finally, suppose that \( U \) is in the kernel of \( Q \). Then \( Q(U) = id_{\mathbb{P}(\mathcal{H})} \). Now select an arbitrary unit vector \( e \) in \( \mathcal{H} \). Then \([Ue] = Q(U)([e]) = [e]\). Since \( U \) is either unitary
or anti-unitary, $U e$ is also a unit vector so there is a unique $z \in \mathbb{C}$ with $|z| = 1$ such that $U e = z e$, that is, $z^{-1} U e = e$. Consequently,

$$Q(z^{-1} U)([e]) = [(z^{-1} U)e] = [e].$$

Notice also that

$$Q(I)([e]) = [I e] = [e].$$

The uniqueness assertion in Wigner’s Theorem then implies that $z^{-1} U = I$ so $U = z I$ as required.

Identifying $Aut(\mathcal{P}(\mathcal{H}))$ with $[\mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^{-}(\mathcal{H})]/S^1$ we provide it with the quotient topology determined by $Q$ and one can show that it thereby acquires the structure of a Hausdorff topological group.

Now let $G$ be an arbitrary Hausdorff topological group. A projective representation of $G$ on the complex, separable Hilbert space $\mathcal{H}$ is a continuous group homomorphism $\rho : G \to Aut(\mathcal{P}(\mathcal{H}))$ of $G$ into the automorphism group $Aut(\mathcal{P}(\mathcal{H}))$. Two projective representation $\rho_j : G \to Aut(\mathcal{P}(\mathcal{H}_j))$, $j = 1, 2$, of $G$ are equivalent if there is an isomorphism $T : \mathcal{P}(\mathcal{H}_1) \to \mathcal{P}(\mathcal{H}_2)$ of $\mathcal{P}(\mathcal{H}_1)$ onto $\mathcal{P}(\mathcal{H}_2)$ such that $\rho_2(g) \circ T = T \circ \rho_1(g)$ for every $g \in G$. Notice that any representation $\tilde{\rho} : G \to \mathcal{U}(\mathcal{H}_1) \sqcup \mathcal{U}^{-}(\mathcal{H})$ of $G$ into $\mathcal{U}(\mathcal{H}_1) \sqcup \mathcal{U}^{-}(\mathcal{H})$ gives rise to projective representation of $G$ by composing with $Q$

$$\rho = Q \circ \tilde{\rho}.$$

It is not the case, however, that every projective representation arises in this way. We will say that $\rho : G \to Aut(\mathcal{P}(\mathcal{H}))$ lifts if there exists a continuous homomorphism $\tilde{\rho} : G \to \mathcal{U}(\mathcal{H}) \sqcup \mathcal{U}^{-}(\mathcal{H})$ for which $\rho = Q \circ \tilde{\rho}; \tilde{\rho}$ is then called a lift of $\rho$. As with any lifting problem for continuous maps the existence of such lifts is a topological issue. For connected, simply connected Lie groups $G$ the obstruction is the second Lie algebra cohomology $H^2(g; \mathbb{R})$.

**Remark 1.3.2.** Lie algebra cohomology can be introduced in a variety of ways. A very thorough discussion together with a number of applications to physics is available in [deAI]. We will not pursue this here since our interest in the general result is limited to the fact that it implies a theorem of Bargmann (Theorem 2.7.1) that we will state, but not prove in Section 2.7. Nevertheless, we will, for the record, formulate the result precisely. The following is Theorem 4, Section 2, of [Simms] and Corollary 3.12 of [VDB].

**Theorem 1.3.3.** Let $G$ be a connected, simply connected Lie group whose Lie algebra $\mathfrak{g}$ satisfies $H^2(\mathfrak{g}; \mathbb{R}) = 0$ and let $\mathcal{H}$ be a complex, separable Hilbert space. Then every projective representation $\rho : G \to Aut(\mathcal{P}(\mathcal{H}))$ of $G$ on $\mathcal{P}(\mathcal{H})$ lifts to a unitary representation $\tilde{\rho} : G \to \mathcal{U}(\mathcal{H})$ of $G$ on $\mathcal{H}$. 


Notice that, under the circumstances described in the Theorem, the lift \( \tilde{\rho} \) actually maps into the unitary subgroup \( \mathcal{U}(H) \) of \( \mathcal{U}(\mathfrak{g}) \cup \mathcal{U}^-(\mathfrak{g}) \).

1.4 Induced Representations

Let \( G \) be a Lie group, \( H \) a closed subgroup of \( G \) and \( \pi : G \to G/H \) the natural projection of \( G \) onto the set of left cosets of \( H \) in \( G \). We supply \( G/H \) with the differentiable structure described in Theorem 1.1.5. Now notice that right multiplication by elements of \( H \) defines a smooth right action of \( H \) on \( G \) that preserves the cosets of \( H \), that is, satisfies

\[
\pi(g \cdot h) = \pi(gh) = [gh] = (gh)H = gH = [g] = \pi(g)
\]

for all \( g \in G \) and all \( h \in H \).

To put this in the proper context we will need to recall a few basic facts about principal bundles and their associated vector bundles. The most authoritative source for this material is [KN1], but one might also consult Chapter 4, Section 5.4, and Section 6.7 of [Nab2].

Let \( P \) and \( X \) be smooth manifolds, \( \pi : P \to X \) a smooth map of \( P \) onto \( X \), and \( H \) a Lie group. We say that \( \pi : P \to X \) has the structure of a smooth, principal \( H \)-bundle if there is a smooth right action \( (p \mapsto h) \in P \times H \mapsto p \cdot h \in P \) of \( H \) on \( P \) such that

1. \( \pi(p \cdot h) = \pi(p) \) for all \( p \in P \) and all \( h \in H \).
2. (Local Triviality) For each \( x_0 \in X \) there exists an open neighborhood \( U \) of \( x_0 \) in \( X \) and a diffeomorphism \( \Psi : \pi^{-1}(U) \to U \times H \) of the form

\[
\Psi(p) = (\pi(p), \psi(p)),
\]

where \( \psi : \pi^{-1}(U) \to H \) satisfies

\[
\psi(p \cdot h) = \psi(p)h
\]

for all \( p \in \pi^{-1}(U) \) and all \( h \in H \).

\( P \) is the bundle space or total space, \( X \) is the base space, \( H \) is the structure group and \( \pi \) is the projection of the principal \( H \)-bundle \( \pi : P \to X \). It follows from local triviality that, for each \( x \in X \), the fiber \( \pi^{-1}(x) \) above \( x \) is a submanifold of \( P \) diffeomorphic to \( H \). One thinks of \( P \) as a family of copies of \( H \) parametrized by the points of \( X \) and glued together topologically in such a way as to achieve local triviality. For a given \( X \) and \( H \) there are generally many ways to do this gluing and it is sometimes possible to classify them (see Section 6.4 of [Nab3]). If, in the local triviality condition, one can take \( U \) to be all of \( X \), then the principal \( H \)-bundle is said to be trivial. One can show that any principal \( H \)-bundle over any \( \mathbb{R}^n \) is necessarily trivial. If \( U \) is an open subset of \( X \), then a smooth map \( s : U \to \pi^{-1}(U) \subseteq P \) for
which \( \pi \circ s = id_U \) is called a (local) section of the principal \( H \)-bundle \( \pi : P \to X \).

If \( U = X \), then \( s \) is called a global section of \( \pi : P \to X \). By local triviality, local sections always exist, but global sections exist if and only if the principal \( H \)-bundle is trivial. Generally, when the context makes the rest clear, it is common to refer to \( \pi : P \to X \), or even just to \( P \), simply as a principal bundle over \( X \).

For the homogeneous manifold \( \pi : G \to G/H \) we take the right action of \( H \) on \( G \) to be right multiplication so that, as we have just seen, \( \pi(g \cdot h) = \pi(g) \). The local triviality condition follows from the existence of local sections (see Theorem 1.1.5).

Indeed, if \( s : U \to \pi^{-1}(U) \) is such a section, then

\[
\pi^{-1}(U) = \bigcup_{[g] \in U} \{ s([g]) \cdot h : h \in H \}
\]

and we can define

\[
\Psi(s([g]) \cdot h) = \left( \pi(s([g]) \cdot h), h \right) = ([g], h).
\]

One shows that \( \Psi \) is a diffeomorphism of \( \pi^{-1}(U) \) onto \( U \times H \) so that \( \pi : G \to G/H \) has the structure of a smooth, principal \( H \)-bundle.

There is a standard procedure (described in detail in Section 6.7 of [Nab2]) for constructing from a smooth, principal \( H \)-bundle \( \pi : P \to X \) and a manifold \( M \) on which \( H \) acts smoothly on the left a smooth “fiber bundle” in which each \( H \)-fiber is replaced by a copy of \( M \). In particular, if \( M \) is a finite-dimensional vector space \( V \) and the left action of \( H \) on \( V \) is given by a smooth representation of \( H \) on \( V \), then the result is a smooth “vector bundle” over \( X \). We will need an analogous construction in which \( V \) is replaced by an infinite-dimensional, complex, separable Hilbert space \( \mathcal{H} \) on which some strongly continuous, unitary representation \( \sigma : H \to \mathcal{U}(\mathcal{H}) \) of \( H \) acts. The procedure is the same as in the finite-dimensional case, but because the representation is only assumed continuous the end result, called a \( C^0 \)-Hilbert bundle, will live in the topological category. We will sketch the construction.

Let \( \pi : P \to X \) be a smooth, principal \( H \)-bundle, \( \mathcal{H} \) a complex, separable Hilbert space, and \( \sigma : H \to \mathcal{U}(\mathcal{H}) \) a strongly continuous, unitary representation of \( H \) on \( \mathcal{H} \). Write \( \sigma(h)(v) = h \cdot v \) for all \( h \in H \) and all \( v \in \mathcal{H} \). On the product space \( P \times \mathcal{H} \) define a continuous right action of \( H \) by

\[
(p, v) \cdot h = (p \cdot h, h^{-1} \cdot v).
\]

(1.4)

This action defines an equivalence relation on \( P \times \mathcal{H} \). Specifically, \( (p_1, v_1) \sim (p_2, v_2) \) if and only if there exists an \( h \in H \) such that \( (p_2, v_2) = (p_1, v_1) \cdot h \). The equivalence class of \( (p, v) \) is denoted \([p, v]\) and the set of all such is denoted \( P \times_\sigma \mathcal{H} \). The natural projection of \( P \times \mathcal{H} \) onto \( P \times_\sigma \mathcal{H} \) is denoted \( q_\sigma : P \times \mathcal{H} \to P \times_\sigma \mathcal{H} \) and defined by \( q_\sigma(p, v) = [p, v] \). We provide \( P \times_\sigma \mathcal{H} \) with the quotient topology determined by \( q_\sigma \) and call it the orbit space of the action (1.4). Next define

\[
\pi_\sigma : P \times_\sigma \mathcal{H} \to X
\]
by \( \pi_\sigma([p, v]) = \pi(p) \). This is well-defined because \( \pi(p \cdot h) = \pi(p) \), surjective because \( \pi \) is surjective, and continuous because its composition with \( q_\sigma \) is continuous. For any \( x \in X \) the fiber of \( \pi_\sigma \) above \( x \) is just \( \pi^{-1}_\sigma(x) = \{ [p, v] : v \in \mathcal{H} \} \), where \( p \) is any fixed point in \( \pi^{-1}(x) \). Each of these fibers can be supplied with the structure of a complex, separable Hilbert space isomorphic to \( \mathcal{H} \) by defining \([p, v_1] + [p, v_2] = [p, v_1 + v_2],\) \( d[p, v] = [p, av] \) and \( \langle [p, v_1], [p, v_2] \rangle = \langle v_1, v_2 \rangle_{\mathcal{H}} \). All of this is easy to check. It takes just a bit more work to show that \( \pi_\sigma : \mathcal{P} \times_\sigma \mathcal{H} \rightarrow X \) satisfies a (topological) local triviality condition, that is, that for each \( x_0 \in X \) there is an open neighborhood \( U \) of \( x_0 \) in \( X \) and a homeomorphism \( \Phi : \pi^{-1}_\sigma(U) \rightarrow U \times \mathcal{H} \) of \( \pi^{-1}_\sigma(U) \) onto \( U \times \mathcal{H} \) of the form \( \Phi([p, v]) = (\pi_\sigma([p, v]), \phi([p, v])) = (\pi(p), \phi([p, v])) \), where the restriction \( \phi|_{\pi^{-1}_\sigma(U)} \) of \( \phi \) to each fiber of \( \pi_\sigma \) is a linear homeomorphism. To see where \( \Phi \) comes from begin with a local trivialization \( \Psi : \pi^{-1}(U) \rightarrow U \times H \) of the principal bundle \( \pi : \mathcal{P} \rightarrow X \). Let \( s : U \rightarrow \pi^{-1}(U) \) be the natural associated section defined by \( s(x) = \Psi^{-1}(x, e) \). Now define \( \Phi^{-1} : U \times \mathcal{H} \rightarrow \pi^{-1}_\sigma(U) \) by \( \Phi^{-1}(x, v) = [s(x), v] \). The \( \Phi \) we are looking for is the inverse of this.

With the structure we have just described \( \pi_\sigma : \mathcal{P} \times_\sigma \mathcal{H} \rightarrow X \) is an example of a \( C^0 \)-Hilbert bundle over \( X \). A (global) section of \( \pi_\sigma : \mathcal{P} \times_\sigma \mathcal{H} \rightarrow X \) is a continuous map \( s : X \rightarrow \mathcal{P} \times_\sigma \mathcal{H} \) with the property that \( \pi_\sigma \circ s = \text{id}_X \). One can show (see Section 6.8 of [Nab2] and Remark 1.4.1 below) that these sections are in one-to-one correspondence with continuous maps \( f : \mathcal{P} \rightarrow \mathcal{H} \) that are equivariant with respect to the given actions of \( H \) on \( \mathcal{P} \) and \( \mathcal{H} \), that is, that satisfy \( f(p \cdot h) = h^{-1} \cdot f(p) = \sigma(h)^{-1}(f(p)) \).

For the time being we will identify continuous sections of \( \pi_\sigma : \mathcal{P} \times_\sigma \mathcal{H} \rightarrow X \) with these continuous, equivariant, \( \mathcal{H} \)-valued maps on \( \mathcal{P} \).

**Remark 1.4.1.** In Section 2.8 we will need to undo this identification and convert the equivariant maps to sections so we will note for the record that the section \( s_f : X \rightarrow \mathcal{P} \times_\sigma \mathcal{H} \) corresponding to the equivariant map \( f : \mathcal{P} \rightarrow \mathcal{H} \) is given by \( s_f(x) = [p, f(p)] \), where \( p \) is any point in \( \pi^{-1}(x) \).

With this review behind us we can return to the problem of describing induced representations. Thus, we let \( G \) be a Lie group, \( H \) a closed subgroup of \( G \), and \( \sigma : H \rightarrow \mathcal{U}(\mathcal{H}) \) a strongly continuous, irreducible, unitary representation of \( H \) on the complex, separable Hilbert space \( \mathcal{H} \). We would like to construct from this data a strongly continuous, irreducible, unitary representation of \( G \) on some complex, separable Hilbert space \( \mathcal{H}^\sigma \). We now know that \( \pi : G \rightarrow G/H \) is a principal bundles over \( G/H \) in which the right action of \( H \) on \( G \) is just right multiplication by elements of \( H \). This, together with \( \sigma : H \rightarrow \mathcal{U}(\mathcal{H}) \), then gives rise to an associated \( C^0 \)-Hilbert bundle \( \pi^\sigma : G \times_\sigma \mathcal{H} \rightarrow G/H \). The Hilbert space \( \mathcal{H}^\sigma \) will be the space of \( L^2 \)-sections of \( \pi^\sigma : G \times_\sigma \mathcal{H} \rightarrow G/H \). Naturally, this requires that we be able to integrate over \( G/H \). Now \( G \), being a locally compact group, has a left-invariant Haar
measure on the Borel sets of $G$ that is unique up to a normalizing factor. However, this measure generally does not give rise to a $G$-invariant measure on $G/H$ and this is what we need. It does give rise to a quasi-invariant measure on $G/H$, meaning that the $G$-action on $G/H$ preserves sets of measure zero, and it turns out that this is enough. Nevertheless, in all of the examples of interest to us $G/H$ will admit a natural $G$-invariant measure so for the remainder of our sketch we will simply assume that $G/H$ admits a measure $\mu$ that is invariant under the $G$-action on $G/H$.

Now the construction of $\mathcal{H}^G$ proceeds as follows. Begin with the set $C^0_0(G, \mathcal{H})$ of continuous maps $f : G \to \mathcal{H}$ satisfying
1. $f(gh) = \sigma(h)^{-1}f(g) \ \forall g \in G \ \forall h \in H$, and
2. $\{[g] \in G/H : g \in \text{supp}(f)\}$ is compact in $G/H$.

Notice that, since $\sigma$ is unitary, $\|f(gh)\|_{\mathcal{H}} = \|f(g)\|_{\mathcal{H}}$ for all $g \in G$ and all $h \in H$ so $\|f(g)\|_{\mathcal{H}}$ depends only on $[g]$. We can therefore define a real-valued function on $G/H$ that we will denote $\|f([g])\|$ by

$$\|f([g])\| = \|f(g)\|_{\mathcal{H}},$$

where $g$ is any element of $[g]$. Integrating with respect to the $G$-invariant measure $\mu$ on $G/H$ we define

$$\|f\| = \left( \int_{G/H} \|f([g])\|^2 d\mu([g]) \right)^{1/2}.$$

This is a norm on $C^0_0(G, \mathcal{H})$ and arises from the inner product

$$\langle f_1, f_2 \rangle = \int_{G/H} \langle f_1([g]), f_2([g]) \rangle d\mu([g]),$$

where $\langle f_1([g]), f_2([g]) \rangle = \langle f_1(g), f_2(g) \rangle_{\mathcal{H}}$; this also depends only on $[g]$. Since the elements of $C^0_0(G, \mathcal{H})$ are continuous, it is not complete with respect to this norm so we take $\mathcal{H}^G$ to be its Hilbert space completion. The elements of this completion can be identified with Borel measurable functions $f : G \to \mathcal{H}$, modulo equality almost everywhere with respect to Haar measure on $G$, that satisfy
1. $f(gh) = \sigma(h)^{-1}f(g)$ for all $h \in H$ and almost all $g \in G$, and
2. $\int_{G/H} \|f([g])\|^2 d\mu([g]) < \infty$.

We can now fulfill our stated objective. Thus, we let $G$ be a Lie group, $H$ a closed subgroup of $G$, and $\sigma : H \to \text{U}(\mathcal{H})$ a strongly continuous, irreducible, unitary representation of $H$ on the complex, separable Hilbert space $\mathcal{H}$. Then the representation of $G$ induced by $H$ and $\sigma$ is denoted

$$\text{Ind}_H^G(\sigma) : G \to \mathcal{U}(\mathcal{H}^G)$$

and acts by left translation on the elements of $\mathcal{H}^G$, that is,
[ \left[ \operatorname{Ind}_H^G(\sigma)(g) \right](g') = f(g^{-1}g').

Exercise 1.4.1. Show that $\operatorname{Ind}_H^G(\sigma)$ is a unitary representation of $G$ on $\mathcal{C}$.

We are most interested in the induced representation when $G$ is given as a semi-direct product and it is to this case that we will turn in the next section.

1.5 Representations of Semi-Direct Products

We begin with some algebraic generalities on semi-direct products of groups. Let $N$ and $H$ be groups and $\theta : H \to \text{Aut}(N)$ a homomorphism of $H$ into the automorphism group of $N$. Then $\theta$ determines a left action of $H$ on $N$ which we will write as

$$h \cdot n = \theta(h)(n)$$

for all $h \in H$ and all $n \in N$.

Exercise 1.5.1. Verify that

$$h_1 \cdot (h_2 \cdot n) = (h_1 h_2) \cdot n$$

and

$$h \cdot (n_1 n_2) = (h \cdot n_1)(h \cdot n_2)$$

for all $h_1, h_2, h \in H$ and all $n_1, n_2, n \in N$.

The semi-direct product of $N$ and $H$ determined by $\theta$ is the group

$$G = N \rtimes_\theta H$$

whose underlying set is $N \times H = \{ (n, h) : n \in N, h \in H \}$ and in which the group operations are defined by

$$(n_1, h_1)(n_2, h_2) = (n_1 h_1 \cdot n_2, h_1 h_2),$$

$$1_G = (1_N, 1_H),$$

$$(n, h)^{-1} = (h^{-1} \cdot n^{-1}, h^{-1}).$$

Exercise 1.5.2. Verify the group axioms and show that the maps

$$n \mapsto (n, 1_H) : N \to N \rtimes_\theta H$$
and

\[ h \mapsto (1_N, h) : H \to N \rtimes_\theta H \]

are embeddings so that we can (and will) identify \( N \) and \( H \) with subgroups of \( N \rtimes_\theta H \). Notice also that, when \( \theta \) is the trivial homomorphism that sends everything to the identity, the semi-direct product reduces to the usual direct product \( N \times H \) of the groups \( N \) and \( H \).

**Remark 1.5.1.** When the action of \( H \) on \( N \) is clear from the context we will often omit the subscript \( \theta \) and write \( N \rtimes_\theta H \) simply as \( N \rtimes H \).

Now notice that

\[
(1_N, h)(n, 1_H)(1_N, h)^{-1} = (1_N(h \cdot n), h1_H)(h^{-1} \cdot 1_N^{-1}, h^{-1}) \\
= (h \cdot n, h)(1_N, h^{-1}) \\
= ((h \cdot n)1_N, hh^{-1}) \\
= (h \cdot n, 1_H) \\
= (\theta(h)(n), 1_H).
\]

If we identify \( N \) and \( H \) with subgroups of \( G = N \rtimes_\theta H \) in the manner described in the previous exercise this simply says that the action of \( H \) on \( N \) is by conjugation in \( G \), that is,

\[ h \cdot n = hnh^{-1}. \]

**Exercise 1.5.3.** Identify \( N \) and \( H \) with subgroups of \( G = N \rtimes_\theta H \) and show that

1. \( N \) is a normal subgroup of \( G \),
2. \( NH = G \), and
3. \( N \cap H = \{1_G\} \).

Show also that the last two conditions imply that every element \( g \) of \( G \) can be written *uniquely* as \( g = nh \) with \( n \in N \) and \( h \in H \).

Turning matters around let us suppose that \( G \) is an arbitrary group and \( N \) and \( H \) are subgroups of \( G \) satisfying the conditions

1. \( N \) is a normal subgroup of \( G \),
2. \( NH = G \), and
3. \( N \cap H = \{1_G\} \).

Then \( G \) is said to be the *internal semi-direct product* of \( N \) and \( H \). One can then define an action of \( H \) on \( N \) by conjugation, that is, define \( \theta : H \to Aut(N) \) by \( \theta(h)(n) = hnh^{-1} \).
Exercise 1.5.4. Show that, in this case, the map \( \phi : N \rtimes_0 H \rightarrow G \) defined by

\[
\phi(n, h) = nh
\]

is an isomorphism. Consequently, if \( G \) is the internal semi-direct product of the subgroups \( N \) and \( H \), then it is also the (external) semi-direct product of the groups \( N \) and \( H \).

The examples of most interest to us will be described in Section 2.4, but in order to see how semi-direct products arise in practice and to ease our way back into the category of Lie groups we will look at the following simpler example.

Example 1.5.1. (The Inhomogeneous Rotation Group) Fix an element \( R \) of \( \text{SO}(3) \) and an \( a \) in \( \mathbb{R}^3 \). Define a mapping \( (a, R) : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) by

\[
x \in \mathbb{R}^3 \rightarrow (a, R)(x) = Rx + a \in \mathbb{R}^3,
\]

where \( Rx \) is the matrix product of \( R \) and the (column) vector \( x \). Thus, \( (a, R) \) rotates by \( R \) and then translates by \( a \) so it is an isometry of \( \mathbb{R}^3 \). The composition of two such mappings is given by

\[
x \rightarrow R_2x + a_2 \rightarrow R_1(R_2x + a_2) + a_1 = (R_1R_2)x + (R_1a_2 + a_1).
\]

Since \( R_1, R_2 \in \text{SO}(3) \) and \( R_1a_2 + a_1 \in \mathbb{R}^3 \), this composition is just

\[
(a_1, R_1) \circ (a_2, R_2) = (R_1a_2 + a_1, R_1R_2)
\]

so this set of mappings is closed under composition. Moreover, \((0, \text{id}_{\mathbb{R}^3})\) is clearly an identity element and every \((a, R)\) has an inverse given by

\[
(a, R)^{-1} = (-R^{-1}a, R^{-1})
\]

so this collection of maps forms a group under composition. This group is the semi-direct product of \( \mathbb{R}^3 \) and \( \text{SO}(3) \) corresponding to the natural action of \( \text{SO}(3) \) on \( \mathbb{R}^3 \). We will denote it \( \text{ISO}(3) \) and refer to it as the inhomogeneous rotation group. Its elements are diffeomorphisms of \( \mathbb{R}^3 \) onto itself and we can think of it as defining a group action on \( \mathbb{R}^3 \).

\[
(a, R) \cdot x = Rx + a
\]

Notice that the maps \( a \rightarrow (a, \text{id}_{\mathbb{R}^3}) \) and \( R \rightarrow (0, R) \) identify \( \mathbb{R}^3 \) and \( \text{SO}(3) \) with subgroups of \( \text{ISO}(3) \) and that \( \mathbb{R}^3 \) is a normal subgroup since it is the kernel of the projection \( (a, R) \rightarrow (0, R) \) and this is a homomorphism (the projection onto \( \mathbb{R}^3 \) is not a homomorphism).

We would like to find an explicit matrix model for \( \text{ISO}(3) \). For this we identify \( \mathbb{R}^3 \) with the subset of \( \mathbb{R}^4 \) consisting of (column) vectors of the form
\[
\begin{pmatrix}
  x^1 \\
  x^2 \\
  x^3 \\
  1
\end{pmatrix}
= \begin{pmatrix}
  x \\
  1
\end{pmatrix}
\]

where \( x = (x^1, x^2, x^3)^T \in \mathbb{R}^3 \). Now consider the set \( G \) of \( 4 \times 4 \) matrices of the form

\[
\begin{pmatrix}
  R & a \\
  0 & 1
\end{pmatrix} = \begin{pmatrix}
  R_1 & R_1^2 & R_1^3 & a^1 \\
  R_2 & R_2^2 & R_3^3 & a^2 \\
  R_3 & R_3^2 & R_3^3 & a^3 \\
  0 & 0 & 0 & 1
\end{pmatrix},
\]

where \( R \in \text{SO}(3) \) and \( a \in \mathbb{R}^3 \). Notice that

\[
\begin{pmatrix}
  R & a \\
  0 & 1
\end{pmatrix} \begin{pmatrix}
  x \\
  1
\end{pmatrix} = \begin{pmatrix}
  Rx + a \\
  1
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
  R_1 & a_1 \\
  0 & 1
\end{pmatrix} \begin{pmatrix}
  R_2 & a_2 \\
  0 & 1
\end{pmatrix} = \begin{pmatrix}
  R_1R_2 & R_1a_2 + a_1 \\
  0 & 1
\end{pmatrix}
\]

so we can identify \( \text{ISO}(3) \) with \( G \) and its action on \( \mathbb{R}^3 \) with matrix multiplication.

\( G \) is a matrix Lie group of dimension 6. Its Lie algebra can be identified with the set of \( 4 \times 4 \) real matrices that arise as velocity vectors to curves in \( G \) through the identity with matrix commutator as bracket. We find a basis for this Lie algebra (otherwise called a set of generators) by noting that if

\[
\alpha_a(t) = \begin{pmatrix}
  \text{id}_{3 \times 3} & t \cdot a \\
  0 & 1
\end{pmatrix},
\]

then

\[
\alpha'_a(0) = \begin{pmatrix}
  0 & a \\
  0 & 0
\end{pmatrix}
\]

and if

\[
\alpha_N(t) = \begin{pmatrix}
  e^N & 0 \\
  0 & 1
\end{pmatrix},
\]

then

\[
\alpha'_N(0) = \begin{pmatrix}
  N & 0 \\
  0 & 0
\end{pmatrix}
\]

(see Theorem A.2.2 for \( N \)). Taking \( a = (1, 0, 0), (0, 1, 0), (0, 0, 1) \) and \( \hat{a} = (1, 0, 0), (0, 1, 0), (0, 0, 1) \) (again, see Theorem A.2.2 for \( \hat{a} \)) we obtain a set of six generators for the Lie algebra \( \text{iso}(3) \) of \( \text{ISO}(3) \) that we will write as follows.

\[
N_1 = \begin{pmatrix}
  0 & 0 & 0 \\
  0 & -1 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 0
\end{pmatrix},
N_2 = \begin{pmatrix}
  0 & 0 & 1 \\
  0 & 0 & 0 \\
  -1 & 0 & 0 \\
  0 & 0 & 0
\end{pmatrix},
N_3 = \begin{pmatrix}
  0 & -1 & 0 \\
  1 & 0 & 0 \\
  0 & 0 & 0 \\
  0 & 0 & 0
\end{pmatrix}
\]
\begin{equation*}
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}, 
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}, 
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\end{equation*}

$N_1, N_2, \text{ and } N_3$ are called the \textit{generators of rotations}, while $P_1, P_2, \text{ and } P_3$ are called the \textit{generators of translations}.

We will write $[A, B] = AB - BA$ for the matrix commutator. Using $\epsilon_{ijk}$ for the Levi-Civita symbol ($1$ if $ijk$ is an even permutation of $1 2 3$, $-1$ if $ijk$ is an odd permutation of $1 2 3$ and $0$ otherwise) we record the following \textit{commutation relations} for these generators, all of which can be verified by simply computing the matrix products.

\begin{align*}
[N_i, N_j] &= \epsilon_{ijk} N_k, \quad i, j = 1, 2, 3 \\
[P_i, P_j] &= 0, \quad i, j = 1, 2, 3 \\
[N_i, P_j] &= \epsilon_{ijk} P_k, \quad i, j = 1, 2, 3
\end{align*}

The following result on semi-direct products of Lie groups is proved in Chapter IV, Section XV, of [Chev].

\textbf{Theorem 1.5.1.} Let $N$ and $H$ be Lie groups. Then, with the compact-open topology, $\text{Aut}(N)$ is also a Lie group and so any continuous group homomorphism $\theta : H \rightarrow \text{Aut}(N)$ is smooth. The corresponding semi-direct product $N \rtimes_\theta H$ is also a Lie group.

Our final objective in this section is to apply the construction of induced representations (Section 1.4) to certain semi-direct products of Lie groups in order to describe the so-called \textit{Mackey machine} for manufacturing all of the irreducible, unitary representations of such groups. Our primary goal is to apply this machine to the Poincaré group and its universal double cover in Section 2.8.

We will consider two Lie groups $N$ and $H$, where $N$ is assumed Abelian, and a continuous (and therefore smooth) homomorphism $\theta : H \rightarrow \text{Aut}(N)$ of $H$ into $\text{Aut}(N)$. As usual, we will write $\theta(h)(n) = h \cdot n$. Then

$$G = N \rtimes_\theta H$$

is also a Lie group. It will be convenient to identify both $N$ and $H$ with closed subgroups of $G$ so that $N$ is a normal subgroup, $G = NH$, $N \cap H = \{e\}$, and the action of $H$ on $N$ is by conjugation

$$h \cdot n = hnh^{-1}.$$

An important role in the construction will be played by the irreducible, unitary representations of $N$. Since $N$ is Abelian these are all 1-dimensional (Corollary
1.2.2) and are described by the \textit{characters} of $N$. We will pause for a moment to provide some background information.

\textbf{Remark 1.5.2.} In our present circumstances $N$ is an Abelian Lie group, but the discussion that follows requires only that $N$ be a locally compact, Hausdorff, Abelian topological group. A \textit{character} of $N$ is a continuous homomorphism $\xi : N \rightarrow S^1$ from $N$ to the group of complex numbers of modulus one. Each of these determines a \textit{unitary representation} $U_\xi$ of $N$ on $\mathbb{C}$ defined by $U_\xi(n)z = \xi(n)z$. The set of all characters of $N$ is denoted $\hat{N}$. Under pointwise multiplication ($(\xi_1\xi_2)(n) = \xi_1(n)\xi_2(n)$) $\hat{N}$ is an Abelian group called the \textit{character group}, or \textit{dual group} of $N$. Regard $\hat{N}$ as a subset of the space $C^0(N, S^1)$ of continuous maps from $N$ to $S^1$ with the compact-open topology and provide it with the subspace topology. $\hat{N}$ thereby becomes a second countable, locally compact, Hausdorff, Abelian topological group. The local compactness is not at all obvious (see page 89 of [Fol2]). If $N_1, \ldots, N_k$ are locally compact, Abelian groups, then so is $N_1 \times \cdots \times N_k$ and the character group of $N_1 \times \cdots \times N_k$ is isomorphic, as a topological group, to $\hat{N}_1 \times \cdots \times \hat{N}_k$ (see Proposition 4.6 of [Fol2]).

An isomorphism from $\hat{N}_1 \times \cdots \times \hat{N}_k$ to the dual group of $N_1 \times \cdots \times N_k$ is given by $(\xi_1, \ldots, \xi_k) \mapsto \xi$, where

$$\xi(n_1, \ldots, n_k) = \xi_1(n_1) \cdots \xi_k(n_k).$$

\textbf{Example 1.5.2.} We will first find all of the characters of the additive group $\mathbb{R}$ and show that $\mathbb{R}$ is isomorphic to $\mathbb{R}$. Thus, we let $\xi : \mathbb{R} \rightarrow S^1$ be a continuous group homomorphism so that $\xi(x_1 + x_2) = \xi(x_1)\xi(x_2)$ and $\xi(0) = 1$. Then there exists an $\epsilon > 0$ such that $\xi([-\epsilon, \epsilon])$ is contained in $\text{Re}(z) > 0$. Since $p \in [-\frac{\pi}{2\epsilon}, \frac{\pi}{2\epsilon}] \Rightarrow pe \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ there is a unique $p \in [-\frac{\pi}{2\epsilon}, \frac{\pi}{2\epsilon}]$ such that

$$\xi(e) = e^{ipe}.$$

Now, $\xi(e) = \xi(\frac{\pi}{2} + \frac{\pi}{2}) = \xi(\frac{\pi}{2})^2$ and therefore

$$\xi\left(\frac{\epsilon}{2}\right) = e^{ip(\epsilon/2)}$$

because $-e^{ip(\epsilon/2)}$ does not have positive real part. Iterating gives

$$\xi\left(\frac{\epsilon}{2^n}\right) = e^{ip(\epsilon/2^n)}, \hspace{1em} n = 0, 1, 2, \ldots$$

Thus, for any $k \in \mathbb{Z}$, $k \frac{\epsilon}{2^n}

$$\xi\left(\frac{k}{2^n} \epsilon\right) = e^{ip(k/2^n)\epsilon}.$$

Since the set of all $\frac{k}{2^n} \epsilon$, $k \in \mathbb{Z}$, $n = 0, 1, 2, \ldots$ is dense in $\mathbb{R}$ and $\xi$ is continuous,
1.5 Representations of Semi-Direct Products

\[ \xi(x) = e^{ipx} \quad \forall x \in \mathbb{R}. \]

**Exercise 1.5.5.** Show that, for each \( \xi \in \hat{\mathbb{R}} \), there exists a unique \( p \) in \( \mathbb{R} \) such that \( \xi(x) = e^{ipx} \) for every \( x \in \mathbb{R} \). **Hint:** Repeat the argument above for any \( \epsilon' \) with \( 0 < \epsilon' < \epsilon \).

Consequently, the map \( \chi : \mathbb{R} \rightarrow \hat{\mathbb{R}} \) defined by \( \chi(p) = e^{ipx} \) is a group isomorphism. All that remains is to show that it is also a homeomorphism and therefore an isomorphism of topological groups. To show that \( \chi \) is continuous it is enough to prove continuity at the identity \( p = 0 \) in \( \mathbb{R} \).

**Exercise 1.5.6.** Show that \( \chi^{-1} : \hat{\mathbb{R}} \rightarrow \mathbb{R} \) is continuous.

We conclude that, as topological groups,

\[ \hat{\mathbb{R}} \cong \mathbb{R}. \]

via the map

\[ p \in \mathbb{R} \mapsto e^{ipx} \in \hat{\mathbb{R}}. \]

Applying the result on products quoted above we find that

\[ \hat{\mathbb{R}}^k \cong \hat{\mathbb{R}} \times \cdots \times \hat{\mathbb{R}} = \hat{\mathbb{R}}^k \cong \mathbb{R}^k \]

and that any element of \( \hat{\mathbb{R}}^k \) can be written in the form

\[ \xi(x^1, \ldots, x^k) = e^{i(p_1 x^1 + \cdots + p_k x^k)}, \]

where \((p_1, \ldots, p_k) \in \mathbb{R}^k\) is unique.

Notice that if \( G = N \rtimes_H H \), then the action of \( H \) on \( N \) induces an action of \( H \) on \( \hat{N} \) as follows. For any \( h \in H \) and any \( \xi \in \hat{N} \) we define \( h \cdot \xi \in \hat{N} \) by
For each orbit be written uniquely as

\[ \forall n \in N. \]

Now fix some \( \xi_0 \in \hat{N} \) and consider the orbit

\[ \mathcal{O}_{\xi_0} = H : \xi_0 = \{h \cdot \xi_0 : h \in H \} \]

of \( \xi_0 \) under the \( H \)-action on \( \hat{N} \). Also let

\[ H_{\xi_0} = \{h \in H : h \cdot \xi_0 = \xi_0\} \]

be the isotropy subgroup of \( \xi_0 \) with respect to this \( H \)-action. This is a closed subgroup of \( H \) and therefore also a Lie group. Physicists refer to \( H_{\xi_0} \) as the little group at \( \xi_0 \). Notice that if \( \xi'_0 \) is another point in \( \mathcal{O}_{\xi_0} \), with, say, \( \xi'_0 = h_0 \cdot \xi_0 \), then the orbit of \( \xi'_0 \) coincides with the orbit of \( \xi_0 \) and the isotropy subgroups \( H_{\xi_0} \) and \( H_{\xi'_0} \) are conjugate (\( H_{\xi_0} = h_0 H_{\xi_0} h_0^{-1} \)) and therefore isomorphic as Lie groups. In particular, any unitary representation \( \sigma \) of \( H_{\xi_0} \) is unitarily equivalent to the unitary representation \( \sigma' \) of \( H_{\xi'_0} \) defined by \( \sigma'(h_0 h_0^{-1}) = \sigma(h_0) \sigma(h) \sigma(h_0)^{-1} \) for \( h \in H_{\xi_0} \). On a fixed orbit in \( \hat{N} \), the little groups all have the same unitary representations, up to unitary equivalence.

**Remark 1.5.3.** There is a technical assumption we will need in order to get the full force of Mackey’s theorem on irreducible, unitary representations of \( G = N \rtimes \sigma H \). We will say that the semi-direct product \( G \) is regular if, for each \( \xi_0 \in \hat{N} \), the orbit \( \mathcal{O}_{\xi_0} \) is locally closed in \( \hat{N} \), meaning that for any \( \xi \in \mathcal{O}_{\xi_0} \) there is an open neighborhood \( V \) of \( \xi \) in \( \hat{N} \) such that \( \mathcal{O}_{\xi_0} \cap V \) is closed in \( V \). Certainly this is the case if all of the orbits are closed in \( \hat{N} \), as they will be for the examples of most interest to us.

Now, suppose we are given a strongly continuous, irreducible, unitary representation \( \sigma \) of the little group \( H_{\xi_0} \) on some separable, complex Hilbert space \( \mathcal{H} \). Assuming that \( H/H_{\xi_0} \) admits an \( H \)-invariant measure (as it will in our examples) we can follow the procedure described in Section 1.4 to produce an induced representation

\[ \text{Ind}_{H_{\xi_0}}^H(\sigma) : H \to \mathcal{U}(\mathcal{H}^\sigma) \]

of \( H \) on the Hilbert space \( \mathcal{H}^\sigma \) of \( L^2 \)-sections of the associated \( C^0 \)-Hilbert bundle. For any \( n \in N \) we let \( U(n) \) be the unitary multiplication operator on \( \mathcal{H}^\sigma \) defined by

\[ [U(n)f](h) = [(h \cdot \xi_0)(n)]f(h) \]

for all \( f \in \mathcal{H}^\sigma \) and \( h \in H \). Finally, recalling that any element \( g \) of \( G = N \rtimes \sigma H \) can be written uniquely as \( g = nh \), where \( n \in N \) and \( h \in H \) we can define \( L_{\mathcal{O}_{\xi_0} \cdot \sigma}^\sigma \) on \( G \) by

\[ L_{\mathcal{O}_{\xi_0} \cdot \sigma}(g) = L_{\mathcal{O}_{\xi_0} \cdot \sigma}(nh) = U(n) \circ \text{Ind}_{H_{\xi_0}}^H(\sigma)(h). \]

For each orbit \( \mathcal{O}_{\xi_0} \) of \( H \) in \( \hat{N} \) and each strongly continuous, irreducible, unitary representation \( \sigma \) of the little group \( H_{\xi_0} \) on some separable, complex Hilbert space \( \mathcal{H} \),
Mackey proves that $L_{O_0,\sigma}$ is a strongly continuous, irreducible, unitary representation of $G = N \rtimes H$ on $\mathcal{H}$. Indeed, much more is true. The following result of Mackey is quite deep and we will simply refer to Theorem 6.24 of [Vara] for the details.

**Theorem 1.5.2.** (Mackey's Theorem) $L_{O_0,\sigma}$ is a strongly continuous, irreducible, unitary representation of $G = N \rtimes H$ on $\mathcal{H}$. If $G$ is a regular semi-direct product, then every strongly continuous, irreducible, unitary representation of $G$ is unitarily equivalent to some $L_{O_0,\sigma}$. Furthermore, $L_{O_{0},\sigma'}$ is unitarily equivalent to $L_{O_{0},\sigma}$ if and only if $O_{\xi_0} = O_{\xi_0}$ and $\sigma'$ is unitarily equivalent to $\sigma$.

Here then is the **Mackey machine** for computing all of the irreducible, unitary representations of the regular semi-direct product $G = NH$ of two Lie groups $N$ and $H$ when $N$ is Abelian. Identify $N$ and $H$ with subgroups of $G$ so that $G = NH$ and the action of $H$ on $N$ is by conjugation.

1. Select an orbit $O$ of the $H$-action on the character group $\hat{N}$ of $N$ and an arbitrary point $\xi_0$ in $O$ so that $O = O_{\xi_0}$.

2. Select an $H$-invariant measure $\mu$ on $H/H_{\xi_0}$, where $H_{\xi_0}$ is the isotropy subgroup of $\xi_0$ in $H$.

*Note:* In general, such a measure need not exist and it is not really necessary for the operation of the Mackey machine, but for the examples of interest to us we will explicitly construct them.

3. Select a strongly continuous, irreducible, unitary representation $\sigma : H_{\xi_0} \to \mathcal{U}(\mathcal{H})$ of the isotropy subgroup $H_{\xi_0}$ of $\xi_0$ on some complex, separable Hilbert space $\mathcal{H}$.

4. Construct the $C^0$-Hilbert bundle associated with $H$, $H_{\xi_0}$, and $\mu$ as follows: $\pi : H \to H/H_{\xi_0}$ is an $H_{\xi_0}$-principal bundle, where the right action of $H_{\xi_0}$ on $H$ is right multiplication. This, together with the representation $\sigma : H_{\xi_0} \to \mathcal{U}(\mathcal{H})$ determines an associated $C^0$-vector bundle

$$\pi_{\sigma} : H \times_{\sigma} \mathcal{H} \to H/H_{\xi_0}$$

over $H/H_{\xi_0}$ with $\mathcal{H}$-fibers.

5. The $L^2$-sections of the vector bundle $\pi_{\sigma} : H \times_{\sigma} \mathcal{H} \to H/H_{\xi_0}$ with respect to the measure $\mu$ form a Hilbert space $\mathcal{H}_{\sigma}$ that can be identified with the space of Borel measurable functions $f : H \to \mathcal{H}$, modulo equality almost everywhere with respect to Haar measure on $H$, satisfying
a. \( f(hh_0) = \sigma(h_0)^{-1}f(h) \forall h_0 \in H_{\xi_0} \) and for almost every \( h \in H \), and

b. \( \int_{H/H_{\xi_0}} \| f([h]) \|^2 \, d\mu([h]) < \infty \), where

\[ \|f([g])\| = \|f(g)\|_{\mathcal{C}}, \]

and \( g \) is any element of \([g]\).

6. The representation of \( H \) induced by \( H_{\xi_0} \) and \( \sigma \) is denoted

\[ \text{Ind}^{H}_{H_{\xi_0}}(\sigma) : H \to \mathcal{U}(\mathcal{C}^\sigma) \]

and acts by left translation on the elements of \( \mathcal{C}^\sigma \), that is,

\[ [\text{Ind}^{H}_{H_{\xi_0}}(\sigma)(h)]f(h') = f(h^{-1}h'). \]

7. For any \( n \in N \) let \( U(n) \) be the unitary multiplication operator on \( \mathcal{C}^\sigma \) defined by

\[ [U(n)f](h) = ([h \cdot \xi_0](n)f(h) = \xi_0(h^{-1} \cdot n)f(h) \]

for all \( f \in \mathcal{C}^\sigma \) and \( h \in H \).

8. Define \( L_{O,\sigma} \) on \( G \) as follows. Write \( g \in G \) uniquely as \( g = nh \), where \( n \in N \) and \( h \in H \). Then

\[ L_{O,\sigma}(g) = L_{O,\sigma}(nh) = U(n) \circ \text{Ind}^{H}_{H_{\xi_0}}(\sigma)(h). \]

Then \( L_{O,\sigma} \) is a strongly continuous, irreducible, unitary representation of \( G \). Up to unitary equivalence this representation is independent of the choice of \( \xi_0 \in O \). \( L_{O',\sigma'} \) is unitarily equivalent to \( L_{O,\sigma} \) if and only if \( O' = O \) and \( \sigma' \) is unitarily equivalent to \( \sigma \). Moreover, every strongly continuous, irreducible, unitary representation of \( G \) is unitarily equivalent to some \( L_{O,\sigma} \).

Needless to say, the machine operates only when the isotropy groups \( H_{\xi_0} \) are sufficiently simple that one can find all of their representations. Fortunately, this is often, although not always, the case. In particular, we will find in Section 2.8 that Mackey’s procedure yields an explicit description of all of the strongly continuous, irreducible, unitary representations of the universal cover of the Poincaré group. These, in turn, give the objects of interest in relativistic quantum mechanics, that is, the projective representations of the Poincaré group.
Chapter 2
Minkowski Spacetime

2.1 Introduction

Quantum Field Theory (QFT) arose from attempts to reconcile quantum mechanics with the special theory of relativity. In this chapter we will attempt to provide the background in relativity required to understand the challenges that such a reconciliation must confront and how the formalism of QFT proposes to deal with them. We will discuss only those aspects of special relativity that are directly relevant to this objective. Our primary source for this material and our primary reference for a more comprehensive introduction to special relativity is [Nab4]. In particular, Section 2.2 is essentially the Introduction to [Nab4].

2.2 Motivation

Minkowski spacetime is generally regarded as the appropriate arena within which to formulate those laws of physics that do not refer specifically to gravitational phenomena. We would like to spend a moment here at the outset briefly examining some of the circumstances that give rise to this belief.

We shall adopt the point of view that the basic problem of science in general is the description of “events” that occur in the physical universe and the analysis of relationships between these events. We use the term “event”, however, in the idealized sense of a “point-event,” that is, a physical occurrence that has no spatial extension and no duration in time. One might picture, for example, an instantaneous collision or explosion, or an “instant” in the history of some material point particle or photon. In this way the existence of a point particle or photon can be represented by a continuous sequence of events called its worldline. We begin then with an abstract set \( \mathcal{M} \) whose elements we call events. We will provide \( \mathcal{M} \) with a mathematical structure that reflects certain simple facts of human experience as well as some rather nontrivial results of experimental physics.
Events are “observed” and we will be particularly interested in a certain class of observers, called admissible, and the means they employ to describe events. Since it is in the nature of our perceptual apparatus that we identify events by their “location in space and time,” we must specify the means by which an observer is to accomplish this in order to be deemed admissible. We begin the process as follows.

Each admissible observer presides over a 3-dimensional, right-handed, Cartesian spatial coordinate system based on an agreed unit of length and relative to which photons propagate rectilinearly in any direction.

A few remarks are in order. First, the expression “presides over” is not to be taken too literally. An observer is in no sense ubiquitous. Indeed, we generally picture the observer as just another material particle residing at the origin of his spatial coordinate system; any information regarding events that occur at other locations must be communicated to him by means we will consider shortly. Second, the restriction on the propagation of photons is a real restriction. The term “straight line” has meaning only relative to a given spatial coordinate system and if, in one such system, light does indeed travel along straight lines, then it certainly will not in another system which, say, rotates relative to the first. Notice, however, that this assumption does not preclude the possibility that two admissible coordinate systems are in relative motion. We shall denote the spatial coordinate systems of observers $O, \hat{O}, \ldots$ by $\Sigma(x^1, x^2, x^3), \hat{\Sigma}(\hat{x}^1, \hat{x}^2, \hat{x}^3), \ldots$

We take it as a fact of human experience that each observer has an innate, intuitive sense of “temporal order” that applies to events which he experiences directly, that is, to events on his worldline. This sense, however, is not quantitative; there is no precise, reliable sense of “equality” for “time intervals.” We remedy this situation by giving him a watch.

Each admissible observer is provided with an ideal, standard clock based on an agreed unit of time with which to provide a quantitative temporal order to the events on his worldline.

Notice that thus far we have assumed only that an observer can assign a time to each event on his worldline. In order for an observer to be able to assign times to arbitrary events we must specify a procedure for the placement and synchronization of clocks throughout his spatial coordinate system. One possibility is simply to mass produce clocks at the origin, synchronize them and then move them to various other points throughout the coordinate system. However, it has been found that moving clocks about has a most undesirable effect upon them. Two identical and very accurate atomic clocks are manufactured in New York and synchronized. One is placed aboard a passenger jet and flown around the world. Upon returning to New York it is found that the two clocks, although they still “tick” at the same rate, are no longer synchronized. The traveling clock lags behind its stay-at-home twin. Strange, indeed, but it is a fact and we shall come to understand the reason for it shortly.
Remark 2.2.1. I didn’t make this up. The experiment was first performed by J.C. Hafele and R.E. Keating in 1971 (see [HK]).

To avoid this difficulty we will ask our admissible observers to build their clocks at the origins of their coordinate systems, transport them to the desired locations, set them down and return to the master clock at the origin. We assume that each observer has stationed an assistant at the location of each transported clock. Now our observer must “communicate” with each assistant, telling him the time at which his clock should be set in order that it be synchronized with the clock at the origin. As a means of communication we choose a signal which seems, among all the possible choices, to be least susceptible to annoying fluctuations in reliability, that is, light signals. To persuade the reader that this is an appropriate choice we will record some of the experimentally documented properties of light signals. First, however, a little experiment. From his location at the origin $O$ an observer $O$ emits a light signal at the instant his clock reads $t_0$. The signal is reflected back to him at a point $P$ and arrives again at $O$ at the instant $t_1$. Assuming there is no delay at $P$ when the signal is bounced back, $O$ will calculate the speed of the signal to be $\text{distance}(O, P)/\sqrt{(t_1 - t_0)}$, where $\text{distance}(O, P)$ is computed from the Cartesian coordinates of $P$ in $\Sigma(x^1, x^2, x^3)$. This technique for measuring the speed of light we call the Fizeau procedure in honor of the gentleman who first carried it out with care.

Remark 2.2.2. Notice that we must bounce the signal back to $O$ since we do not yet have a clock at $P$ that is synchronized with that at $O$.

For each admissible observer the speed of light in vacuo as determined by the Fizeau procedure is independent of when the experiment is performed, the arrangement of the apparatus (that is, the choice of $P$), the frequency (energy) of the signal and, moreover, has the same numerical value $c$ (approximately $3.0 \times 10^8$ meters per second) for all such observers.

Here we have the conclusions of numerous experiments performed over the years, most notably those first performed by Michelson-Morley and Kennedy-Thorndike (see Exercises 33 and 34 of [TW] for a discussion of these experiments). The results may seem odd. Why is a photon so unlike an electron (or a baseball) whose speed certainly will not have the same numerical value for two observers in relative motion? Nevertheless, they are incontestable facts of nature and we must deal with them. We will exploit these rather remarkable properties of light signals immediately by asking all of our observers to multiply each of their time readings by the constant $c$ and thereby measure time in units of distance (light travel time). For example, one meter of time is the amount of time required by a light signal to travel one meter in vacuo. With these units, all speeds are dimensionless and $c = 1$. Such time readings for observers $\hat{O}, \hat{O}, \ldots$ will be denoted $x^0(= ct)$, $\hat{x}^0(= c\hat{t})$, $\ldots$.

Now we provide each of our admissible observers with a system of synchronized clocks in the following way. At each point $P$ of his spatial coordinate system, place
Consequently, the coordinate transformation map and coordinates in following three basic types.  

Proved the remarkable fact that any causal automorphism is a composition of the on the cones. Furthermore, \((2.1)\) onto the cone \((2.2)\) and satisfy \((\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3)\). A useful intuitive visualization of such a reference frame is a lattice work of spatial coordinate lines with, at each lattice point, a clock and an assistant whose task it is to record locations and times for events occurring in his immediate vicinity; the data can later be collected for analysis by the observer.

How are the \(\hat{S}\)-coordinates of an event related to the \(S\)-coordinates? That is, what can be said about the mapping \(F : \mathbb{R}^4 \rightarrow \mathbb{R}^4\) defined by \(F(x^0, x^1, x^2, x^3) = (\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3)\)? Certainly, it must be one-to-one and onto. Indeed, \(F^{-1} : \mathbb{R}^4 \rightarrow \mathbb{R}^4\) must be the coordinate transformation from hatted to unhatted coordinates. To say more we require a seemingly innocuous \textit{Causality Assumption}.

\[
\text{Any two admissible observers agree on the temporal order of any two events on the worldline of a photon, that is, if two such events have coordinates } (x^0, x^1, x^2, x^3) \text{ and } (\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) \text{ in } S \text{ and } (\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) \text{ and } (\hat{\hat{x}}^0, \hat{\hat{x}}^1, \hat{\hat{x}}^2, \hat{\hat{x}}^3) \text{ in } \hat{S}, \text{ then } \Delta x^0 = x^0 - x_0^0 \text{ and } \Delta \hat{x}^0 = \hat{x}^0 - \hat{x}_0^0 \text{ have the same sign.}
\]

Notice that we do not prejudge the issue by assuming that \(\Delta x^0\) and \(\Delta \hat{x}^0\) are equal, but only that they have the same sign, that is, that \(\hat{S}\) and \(\hat{\hat{S}}\) agree as to which of the events occurred first. Thus, \(F\) preserves order in the 0th-coordinate, at least for events that lie on the worldline of a photon. How are two such events related? Since photons propagate rectilinearly with speed 1 in any admissible frame of reference, two events on the worldline of a photon must have coordinates in \(S\) that satisfy

\[
(x^0 - x_0^0)^2 - (x^1 - x_0^1)^2 - (x^2 - x_0^2)^2 - (x^3 - x_0^3)^2 = 0 \tag{2.1}
\]

and coordinates in \(\hat{S}\) that satisfy

\[
(\hat{x}^0 - \hat{x}_0^0)^2 - (\hat{x}^1 - \hat{x}_0^1)^2 - (\hat{x}^2 - \hat{x}_0^2)^2 - (\hat{x}^3 - \hat{x}_0^3)^2 = 0. \tag{2.2}
\]

Consequently, the coordinate transformation map \(F : \mathbb{R}^4 \rightarrow \mathbb{R}^4\) must carry the cone (2.1) onto the cone (2.2) and satisfy

\[
x^0 > x_0^0 \Leftrightarrow x^0 > x_0^0 \tag{2.3}
\]

on the cones. Furthermore, \(F^{-1} : \mathbb{R}^4 \rightarrow \mathbb{R}^4\) carries (2.2) onto (2.1) and satisfies (2.3). In 1964, Zeeman [Z] called such a mapping \(F\) a \textit{causal automorphism} and proved the remarkable fact that any causal automorphism is a composition of the following three basic types.

1. Translations: \(\hat{x}^a = x^a + a^a, a = 0, 1, 2, 3\), for some constants \(a^a, a = 0, 1, 2, 3\).
2. Positive Scalar Multiples: \(\hat{x}^a = k x^a, a = 0, 1, 2, 3\), for some positive constant \(k\).
3. Linear transformations

\[ \hat{x}^\alpha = \Lambda^\alpha_{\beta} x^\beta, \quad \alpha = 0, 1, 2, 3, \quad (2.4) \]

where the matrix \( \Lambda = (\Lambda^\alpha_{\beta})_{\alpha,\beta=0,1,2,3} \) satisfies \( \Lambda^T \eta \Lambda = \eta \), where \( T \) means transpose, \( \eta = (\eta^\alpha_{\beta})_{\alpha,\beta=0,1,2,3} \) is the matrix

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix},
\]

and \( \Lambda^0_0 \geq 1. \)

Remark 2.2.3. Notice that it is not even assumed at the outset that \( F \) is continuous (much less affine). A proof of Zeeman’s Theorem is available either in [Z] or in Section 1.6 of [Nab4]. We should point out that this result was actually proved in 1950 by Alexandrov [Alex], but the paper was in Russian and, sadly, was not widely known in the West.

Since two frames of reference related by a mapping of type (2) differ only by a rather trivial and unnecessary change of scale, we shall banish them from further consideration. In some circumstances one can adopt a similar attitude toward mappings of type (1). The constants \( a^\alpha, \alpha = 0, 1, 2, 3 \), can be regarded as the \( \hat{S} \)-coordinates of \( S \)’s spacetime origin and we may request that all of our observers cooperate to the extent that they select a common event to act as origin, in which case \( a^\alpha = 0, \alpha = 0, 1, 2, 3 \). This is particularly useful when one is interested in purely geometrical questions. On the other hand, in classical mechanics translations in space and time are important symmetries with important conservation laws (momentum and energy) so, when the issue is dynamics, translations will play a fundamental role (see Sections A.2 and A.3).

The (linear) coordinate transformations of type (3), which we will soon christen “orthochronous Lorentz transformations”, contain all of the novel kinematic features of special relativity (“time dilation”, “length contraction”, etc.). We will find that these are precisely the linear maps that leave invariant the quadratic form \((x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2\) (analogous to orthogonal transformations of \( \mathbb{R}^3 \), which leave invariant the usual squared length \( x^2 + y^2 + z^2 \)) and preserve “time orientation” in the sense described in our Causality Assumption. We will see that any such \( \Lambda \) has determinant \( \pm 1 \). Those with \( \det \Lambda = 1 \) are called “proper” and the collection of all proper, orthochronous Lorentz transformations is denoted \( \mathcal{L}^+ \). We will show that \( \mathcal{L}^+ \) forms a group under matrix multiplication (that is, composition). The group generated by \( \mathcal{L}^+ \) and the translations of \( \mathbb{R}^4 \) in (1) is called the Poincaré group and is denoted \( \mathcal{P}^+ \).

The mathematical structure that appears to be emerging for \( \mathcal{M} \) is that of a 4-dimensional real vector space with a distinguished quadratic form and a group of
transformations that preserve the quadratic form. Admissible observers coordinatize \( \mathcal{M} \) and these coordinates are related by elements of the group.

With this we conclude our attempt at motivation for the definitions to follow in Section 2.3. There is, however, one more item on the agenda of our introductory remarks. It is the cornerstone upon which the special theory of relativity is built.

\[ \text{The Relativity Principle: All admissible frames of reference are completely equivalent for the formulation of the laws of physics.} \]

You will object that this is rather vague and we will not dispute the point. One could try to be more precise about what “completely equivalent” means and what “laws of physics” we have in mind, but this would, in some sense, be wrongheaded. It is most profitable to think of the Relativity Principle primarily as a heuristic principle asserting that there are no “distinguished” admissible observers, that is, that none can claim to have a privileged view of the universe. In particular, no such observer can claim to be “at rest” while the others are moving; they are all simply in relative motion. Admissible observers can disagree about some rather startling things (for example, whether or not two given events are “simultaneous”) and the Relativity Principle will prohibit us from preferring the judgement of one to any of the others. Although we will not dwell on the experimental evidence in favor of the Relativity Principle, it should be observed that its roots lie in such commonplace observations as the fact that a passenger in a (smooth, quiet) airplane traveling at constant groundspeed in a straight line cannot “feel” his motion relative to the earth; no physical effects are apparent in the plane that would serve to distinguish it from the (quasi-) admissible frame rigidly attached to the earth.

Our task then is to study the geometry and physics of these “admissible frames of reference.” Before embarking on such a study, however, it is only fair to concede that, in fact, no such thing exists. As with any intellectual construct with which we attempt to model the physical universe, the notion of an admissible frame of reference is an idealization, a rather fanciful generalization of circumstances which, to some degree of accuracy, we encounter in the world. In particular, it has been found that the existence of gravitational fields imposes severe restrictions on the “extent” (both in space and in time) of an admissible frame (for more on this see Section 4.2 of [Nab4]). Knowing this we intentionally avoid the difficulty by restricting our attention to situations in which the effects of gravity are “negligible.”

### 2.3 Geometrical Structure of \( \mathcal{M} \)

\( \text{Minkowski spacetime} \) is a 4-dimensional real vector space \( \mathcal{M} \) on which is defined a nondegenerate, symmetric, bilinear form

\[ \langle \cdot, \cdot \rangle_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \]

of index 3. This means that there exists a basis \( \{e_0, e_1, e_2, e_3\} \) for \( \mathcal{M} \) with
2.3 Geometrical Structure of $\mathcal{M}$

\[
\langle e_\alpha, e_\beta \rangle_\mathcal{M} = \eta_{\alpha\beta} = \begin{cases} 
0, & \text{if } \alpha \neq \beta \\
1, & \text{if } \alpha = \beta = 0 \\
-1, & \text{if } \alpha = \beta = 1, 2, 3
\end{cases}
\]

$\langle , \rangle_\mathcal{M}$ is called the Lorentz inner product or Minkowski inner product on $\mathcal{M}$ and any such basis is said to be $\mathcal{M}$-orthonormal, or simply orthonormal.

Remark 2.3.1. Unless some confusion is likely to arise we will tend to write this and any other inner product that is likely to arise simply as $\langle , \rangle$.

The quadratic form corresponding to $\langle , \rangle$ is the map $Q : \mathcal{M} \rightarrow \mathbb{R}$ defined by

\[Q(x) = \langle x, x \rangle\]

for all $x \in \mathcal{M}$. The points in $\mathcal{M}$ are called events.

Remark 2.3.2. As is customary in linear algebra we will have no qualms about blurring the distinction between a “point” in $\mathcal{M}$ and a “vector” in $\mathcal{M}$ since the context invariably makes clear which geometrical picture we have in mind. If one has moral objections to this the appropriate course is to define Minkowski spacetime as an affine space containing the “points” and distinguish it from the corresponding vector space of “displacement vectors” determined by two “points” (a “tip” and a “tail”).

An $x \in \mathcal{M}$ is said to be spacelike, timelike, or null if $Q(x)$ is $< 0, > 0$, or $= 0$, respectively. We introduce a $4 \times 4$ matrix $\eta$ defined by

\[
\eta = (\eta_{\alpha\beta})_{\alpha,\beta=0,1,2,3} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\]

The inverse of $\eta$ is the same matrix, but we will denote its entries by $\eta^{\alpha\beta}$.

\[
\eta^{-1} = (\eta^{\alpha\beta})_{\alpha,\beta=0,1,2,3} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\]

Writing, with the summation convention, $x = x^\alpha e_\alpha$ we define $x_\beta = \eta_{\alpha\beta}x^\alpha$ for $\beta = 0, 1, 2, 3$. Then $x^\alpha = \eta^{\alpha\beta}x_\beta$ and

\[
\langle x, y \rangle = \langle x^\alpha e_\alpha, y^\beta e_\beta \rangle = \eta_{\alpha\beta}x^\alpha y^\beta = x^\alpha y_\alpha = x_\beta y^\beta.
\]

The basic geometrical structure of $\mathcal{M}$ is discussed in considerable detail in [Nab4]. We will now summarize some of its most important features and then turn to a more detailed discussion of those items that we will specifically call upon in the
Two vectors $x$ and $y$ in $M$ are said to be $M$-orthogonal, or simply orthogonal if $\langle x, y \rangle = 0$. The following results are Theorem 1.2.1 and Corollary 1.3.2 of [Nab4].

**Theorem 2.3.1.** Two nonzero null vectors $x$ and $y$ in $M$ are orthogonal if and only if they are parallel, that is, if and only if there is a $t \neq 0$ in $\mathbb{R}$ such that $y = tx$.

**Theorem 2.3.2.** If a nonzero vector in $M$ is orthogonal to a timelike vector, then it must be spacelike.

Theorem 2.3.2 follows from our next result which provides more detailed information; it is Theorem 1.3.1 of [Nab4].

**Theorem 2.3.3.** Suppose $x$ is timelike and $y$ is either timelike or null and nonzero. Let $\{e_a\}_{a=0,1,2,3}$ be an orthonormal basis for $M$ with $x = x^a e_a$ and $y = y^a e_a$. Then either

1. $x^0 y^0 > 0$, in which case $\langle x, y \rangle > 0$, or
2. $x^0 y^0 < 0$, in which case $\langle x, y \rangle < 0$.

With this we can define an equivalence relation $\sim$ on the collection $\mathcal{T}$ of all timelike vectors in $M$ as follows. For $x, y \in \mathcal{T}$,

$$x \sim y \iff \langle x, y \rangle > 0.$$ 

In this case we say that $x$ and $y$ have the same time orientation. The equivalence relation has precisely two equivalence classes. We arbitrarily select one of these, denote it $\mathcal{T}^+$, and call its elements future directed. The other equivalence class is denoted $\mathcal{T}^-$ and its elements are called past directed. $\mathcal{T}^+$ and $\mathcal{T}^-$ are cones in $M$, that is, if $x$ and $y$ are in $\mathcal{T}^+$ (respectively, $\mathcal{T}^-$) and if $r$ is a positive real number, then $rx$ and $x + y$ are in $\mathcal{T}^+$ (respectively, $\mathcal{T}^-$). One can extend this distinction to nonzero null vectors $n$ by noting that $\langle x, n \rangle$ has the same sign for all $x \in \mathcal{T}^+$. Thus, we can say that $n$ is future directed if $\langle x, n \rangle > 0$ for all $x \in \mathcal{T}^+$ and past directed otherwise.

For each $x_0$ in $M$ we define the time cone $\mathcal{C}_T(x_0)$, future time cone $\mathcal{C}^+_T(x_0)$, and past time cone $\mathcal{C}^-_T(x_0)$ at $x_0$ by

$$\mathcal{C}_T(x_0) = \{ x \in M : x - x_0 \in \mathcal{T} \},$$

$$\mathcal{C}^+_T(x_0) = \{ x \in M : x - x_0 \in \mathcal{T}^+ \},$$

$$\mathcal{C}^-_T(x_0) = \{ x \in M : x - x_0 \in \mathcal{T}^- \}.$$ 

Similarly, the null cone $\mathcal{C}_N(x_0)$, future null cone $\mathcal{C}^+_N(x_0)$, and past null cone $\mathcal{C}^-_N(x_0)$ at $x_0$ are given by
2.3 Geometrical Structure of $\mathcal{M}$

\[ c_\mathcal{M}(x_0) = \{ x \in \mathcal{M} : x - x_0 \text{ is null} \} \]
\[ c_\mathcal{M}^+(x_0) = \{ x \in \mathcal{M} : x - x_0 \text{ is null and future directed} \} \]
\[ c_\mathcal{M}^-(x_0) = \{ x \in \mathcal{M} : x - x_0 \text{ is null and past directed} \} . \]

There is no notion of time orientation for spacelike vectors and the set of spacelike vectors is not a cone so, at each $x_0 \in \mathcal{M}$, one simply defines

\[ c(x_0) = \{ x \in \mathcal{M} : x - x_0 \text{ is spacelike} \} \]

and calls it elsewhere.

With the choice of $\mathcal{T}^+$ we have provided $\mathcal{M}$ with what is called a time orientation. Independently of this we will also select some orientation for the vector space $\mathcal{M}$, that is, some equivalence class of ordered bases for $\mathcal{M}$. Henceforth we will consider only orthonormal bases $\{e_0, e_1, e_2, e_3\}$ for $\mathcal{M}$ that are in the chosen orientation class and for which $e_0$ is timelike and future directed. Such a basis is called an admissible basis for $\mathcal{M}$. The coordinates $(x^0, x^1, x^2, x^3)$ of an event $x$ in $\mathcal{M}$ relative to such a basis are identified with the spatial $(x^1, x^2, x^3)$ and time $(x^0)$ coordinates of the event provided by some admissible observer (see Section 2.2). The choice of such a basis identifies the vector space $\mathcal{M}$ with the vector space $\mathbb{R}^4$ and transfers the Lorentz inner product to $\mathbb{R}^4$. To emphasize the signature of the Lorentz inner product we will write this copy of $\mathbb{R}^4$ as $\mathbb{R}^{1,3}$.

Although the inner product is different, the topology and differentiable structure of $\mathbb{R}^{1,3}$ are taken to be the usual ones of $\mathbb{R}^4$.

**Remark 2.3.3.** It is not unreasonable to argue that the Euclidean topology for $\mathbb{R}^{1,3}$ does not make a great deal of physical sense since the Euclidean inner product has no invariant physical interpretation for all admissible observers. Alternative topologies have been proposed and one of them is discussed in some detail in Appendix A of [Nab4]. Even if one acquiesces to the choice of the Euclidean topology, the differentiable structure is not determined. Deep results of Michael Freedman on topological 4-manifolds and Simon Donaldson on smooth 4-manifolds combine to show that $\mathbb{R}^4$ admits (many) non-diffeomorphic differentiable structures. This cannot occur for any $\mathbb{R}^n$ of dimension $n \neq 4$.

The linear subspace spanned by $e^0$ is called the time axis of the corresponding admissible observer and its orthogonal complement

\[ (e^0)^\perp = \{ x \in \mathcal{M} : \langle e^0, x \rangle = 0 \} = \text{Span}\{e_1, e_2, e_3\} \]

is the observer’s space. We orient $(e^0)^\perp$ by decreeing that $\{e^1, e^2, e^3\}$ is an oriented basis. We will feel free to view $\text{Span}\{e^0\}$ and $\text{Span}\{e^1, e^2, e^3\}$ either in $\mathcal{M}$ or in $\mathbb{R}^{1,3}$.

If $\{e_\alpha\}_{\alpha=0,1,2,3}$ and $\{\tilde{e}_\alpha\}_{\alpha=0,1,2,3}$ are two orthonormal bases for $\mathcal{M}$, then there exists a unique linear transformation $L : \mathcal{M} \to \mathcal{M}$ such that $Le_\alpha = \tilde{e}_\alpha$ and this satisfies
\[ \langle Lx, Ly \rangle = \langle x, y \rangle \]

for all \( x, y \in \mathcal{M} \). Such a linear transformation is called an \( \mathcal{M} \)-orthogonal transformation, or simply an orthogonal transformation. If \( x \in \mathcal{M} \) and if we write \( x = x^\alpha \hat{e}_\alpha \) and \( x = \tilde{x}^\alpha \tilde{\hat{e}}_\alpha \), then \( (x^\alpha) \) and \( (\tilde{x}^\alpha) \) are interpreted as the spacetime coordinates of the event \( x \) in the two frames of reference corresponding to \( \{e_\alpha\}_{\alpha=0,1,2,3} \) and \( \{\tilde{e}_\alpha\}_{\alpha=0,1,2,3} \). We are interested in the coordinate transformation relating these two sets of coordinates. For this we write

\[ e_\gamma = A^\alpha_\gamma \hat{e}_\alpha, \quad \gamma = 0, 1, 2, 3 \]

for some real numbers \( A^\alpha_\gamma, \alpha, \gamma = 0, 1, 2, 3 \). Then the orthogonality conditions \( \langle e_\gamma, e_\delta \rangle = \eta_{\gamma\delta} \) can be written

\[ A^\alpha_\gamma A^\beta_\delta \eta_{\alpha\beta} = \eta_{\gamma\delta}, \quad \gamma, \delta = 0, 1, 2, 3, \tag{2.5} \]

or, equivalently,

\[ A^\alpha_\gamma A^\beta_\delta \eta_{\gamma\delta} = \eta^{\beta\delta}, \quad \alpha, \beta = 0, 1, 2, 3. \tag{2.6} \]

We introduce the \( 4 \times 4 \) matrix

\[
A = (A^\alpha_\beta) = \begin{pmatrix}
A^0_0 & A^0_1 & A^0_2 & A^0_3 \\
A^1_0 & A^1_1 & A^1_2 & A^1_3 \\
A^2_0 & A^2_1 & A^2_2 & A^2_3 \\
A^3_0 & A^3_1 & A^3_2 & A^3_3
\end{pmatrix}
\]

Then the orthogonality conditions (2.5) or, equivalently, (2.6) can be written

\[ A^T \eta A = \eta, \tag{2.7} \]

where “\( T \)” means “transpose”. The coordinate transformation from unhatted to hatted coordinates is just the matrix product

\[
\begin{pmatrix}
x^0 \\
x^1 \\
x^2 \\
x^3
\end{pmatrix} = \begin{pmatrix}
A^0_0 & A^0_1 & A^0_2 & A^0_3 \\
A^1_0 & A^1_1 & A^1_2 & A^1_3 \\
A^2_0 & A^2_1 & A^2_2 & A^2_3 \\
A^3_0 & A^3_1 & A^3_2 & A^3_3
\end{pmatrix}
\begin{pmatrix}
x^0 \\
x^1 \\
x^2 \\
x^3
\end{pmatrix}
\]

or, more concisely,

\[ \hat{x}^\alpha = A^\alpha_\beta x^\beta, \quad \alpha = 0, 1, 2, 3. \tag{2.8} \]

Any real, \( 4 \times 4 \) matrix \( A \) satisfying (2.7) is called a (general) Lorentz transformation. Any such \( A \) satisfies \( \det A = \pm 1 \) and either \( A^0_0 \geq 1 \) or \( A^0_0 \leq -1 \). To ensure that the basis \( \{\tilde{e}_\alpha\}_{\alpha=0,1,2,3} \) has the proper orientation we will consider only Lorentz transformations \( A \) that are proper, that is, satisfy
2.3 Geometrical Structure of $M$

\[ \det \Lambda = 1. \]

To ensure that $\hat{e}^0$ is future directed we assume also that $\Lambda$ is orthochronous, that is, satisfies

\[ \Lambda^0_0 \geq 1. \]

**Remark 2.3.4.** Orthochronous Lorentz transformations actually preserve the time orientation (future directed or past directed) of all timelike and nonzero null vectors (see Theorem 1.3.3 of [Nab4]).

The set of all proper, orthochronous Lorentz transformations is denoted $\mathcal{L}^+_\alpha$.

**Exercise 2.3.1.** Show that $\mathcal{L}^+_\alpha$ is a group under matrix multiplication. *Hint:* Show that any (general) Lorentz transformation $\Lambda$ has an inverse given by

\[ \Lambda^{-1} = \eta \Lambda^T \eta. \]

We will denote the entries of the matrix $\Lambda^{-1}$ by $\Lambda^\alpha_\beta$, where $\alpha$ labels the column and $\beta$ labels the row. Thus,

\[ \Lambda^{-1} = \begin{pmatrix}
  A_0^0 & A_1^0 & A_2^0 & A_3^0 \\
  A_0^1 & A_1^1 & A_2^1 & A_3^1 \\
  A_0^2 & A_1^2 & A_2^2 & A_3^2 \\
  A_0^3 & A_1^3 & A_2^3 & A_3^3
\end{pmatrix} = \begin{pmatrix}
  A^0_0 & -A^0_1 & -A^0_2 & -A^0_3 \\
  -A^1_0 & A^1_1 & A^1_2 & A^1_3 \\
  -A^2_0 & A^2_1 & A^2_2 & A^2_3 \\
  -A^3_0 & A^3_1 & A^3_2 & A^3_3
\end{pmatrix} \]

**Exercise 2.3.2.** Prove each of the following.

1. $\Lambda^\alpha_\beta = \eta_{\epsilon \gamma} \eta^{\beta \delta} A^\gamma_\delta$, $\alpha, \beta = 0, 1, 2, 3$
2. $\Lambda^\alpha_\beta = \eta^{\epsilon \gamma} \eta_{\delta \delta} A^\gamma_\delta$, $\alpha, \beta = 0, 1, 2, 3$
3. $A^\alpha_\gamma \Lambda^\gamma_\delta \eta^\delta_\beta = \eta^\gamma_\beta$, $\gamma, \delta = 0, 1, 2, 3$
4. $A^\alpha_\gamma \Lambda^\gamma_\delta \eta_{\epsilon \delta} = \eta_{\epsilon \beta}$, $\alpha, \beta = 0, 1, 2, 3$

**Exercise 2.3.3.** Show that, if $\Lambda \in \mathcal{L}^+_\alpha$, then $(\Lambda^{-1})^T \in \mathcal{L}^+_\alpha$.

$\mathcal{L}^+_\alpha$ has an important subgroup $\mathcal{R}$ index $\mathcal{R}$ consisting of those $R = (R^\alpha_\beta)_{\alpha, \beta = 0, 1, 2, 3}$ of the form

\[ R = \begin{pmatrix}
  1 & 0 & 0 & 0 \\
  0 & R^1_1 & R^1_2 & R^1_3 \\
  0 & R^2_1 & R^2_2 & R^2_3 \\
  0 & R^3_1 & R^3_2 & R^3_3
\end{pmatrix}. \]
where \((R^a_b)_{a,b=1.2.3}\) is an element of the rotation group SO(3). The coordinate transformation associated with \(R\) corresponds physically to a rotation of the spatial coordinate axes within a given frame of reference. \(R\) is called the rotation subgroup of \(L^+_0\). The following is Lemma 1.3.4 of [Nab4].

**Theorem 2.3.4.** Let \(\Lambda = (\Lambda^a_b)_{a,b=0,1,2,3}\) be a proper, orthochronous Lorentz transformation. Then the following are equivalent.

1. \(\Lambda\) is a rotation in \(L^+_0\).
2. \(\Lambda^0_1 = \Lambda^0_2 = \Lambda^0_3 = 0\)
3. \(\Lambda^1_0 = \Lambda^2_0 = \Lambda^3_0 = 0\)
4. \(\Lambda^0_0 = 1\)

**Exercise 2.3.4.** Let \(\mu\) and \(\nu\) be real numbers with \(\nu > 0\) and \(\nu^2 - \mu^2 = 1\). Show that

\[
\Lambda_{\mu \nu} = \begin{pmatrix}
\nu & 0 & 0 \\
\mu & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

is in \(L^+_0\).

Taking \(\nu = \cosh \theta\) and \(\mu = \sinh \theta\) for some real number \(\theta > 0\) in Exercise 2.3.4 one obtains what is called a boost in the \(x^1\)-direction.

\[
\Lambda_1(\theta) = \begin{pmatrix}
\cosh \theta & \sinh \theta & 0 & 0 \\
\sinh \theta & \cosh \theta & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

The physical interpretation of the corresponding coordinate transformation is discussed in some detail on pages 21-27 of [Nab4]. It describes the relationship between the spacetime coordinates in two admissible frames of reference \(\tilde{S}\) and \(\hat{S}\) for which the spatial coordinate axes initially coincide and remain parallel, but for which those of \(\hat{S}\) move in the positive direction along the common \(x^1\), \(\hat{x}^1\)-axis with speed \(\beta = \tanh \theta\). One can define boosts in the \(x^2\)- and \(x^3\)-directions in an entirely analogous way. Their matrices are

\[
\Lambda_2(\theta) = \begin{pmatrix}
\cosh \theta & 0 & \sinh \theta & 0 \\
0 & 1 & 0 & 0 \\
\sinh \theta & 0 & \cosh \theta & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}, \quad \Lambda_3(\theta) = \begin{pmatrix}
\cosh \theta & 0 & 0 & \sinh \theta \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
\sinh \theta & 0 & 0 & \cosh \theta
\end{pmatrix}.
\]
Remark 2.3.5. These boosts along the spatial coordinate axes of a given frame of reference are generally called *special Lorentz transformations* and we have written them in what is called *hyperbolic form*. They are more traditionally written in terms of the relative velocity \( \beta = \tanh \theta \) of the two frames and \( \gamma = (1 - \beta^2)^{-1/2} \) in which case \( \cosh \theta = \gamma \) and \( \sinh \theta = \beta \gamma \).

Boosts in the \( x^1 \)-direction and rotations suffice to describe all of the elements of the proper, orthochronous Lorentz group. The following is Theorem 1.3.5 of [Nab4].

**Theorem 2.3.5.** Let \( \Lambda \) be a proper, orthochronous Lorentz transformation. Then there exists a real number \( \theta \) and two rotations \( R_1 \) and \( R_2 \) in \( \mathbb{R} \) such that

\[
\Lambda = R_1 \Lambda _1(\theta)R_2.
\]

The Theorem suggests that the boosts \( \Lambda _1(\theta) \) contain a great deal of the kinematic information contained in \( \mathcal{L}_+ \) and this is, indeed, the case. All of the well-known kinematic effects of special relativity such as the relativity of simultaneity, time dilation, length contraction, the relativistic addition of velocities formula, and the quite inappropriately named twin paradox are easily derived from the properties of the coordinate transformation corresponding to \( \Lambda _1(\theta) \). This, however, is not really our business here so, for most of this, we will simply refer to the rather detailed discussions in [Nab4] (particularly pages 29-42). We will, however, need some additional information about timelike vectors and curves that is closely related these phenomena and we will conclude this section with a brief synopsis of what we require.

If \( v \in \mathcal{M} \) is a timelike vector we define its *duration* \( \tau(v) \) by \( \tau(v) = \sqrt{Q(v)} = \sqrt{\langle v, v \rangle} \). If \( v = x - x_0 \) is the displacement vector between two events \( x \) and \( x_0 \) in \( \mathcal{M} \), then it is always possible to find an admissible basis in which \( x_0 \) and \( x \) occur at the same spatial location, one after the other, and \( \tau(x - x_0) \) is interpreted physically as the time separation of \( x \) and \( x_0 \) in any such frame (see pages 42-43 of [Nab4]). In this case \( \tau(x - x_0) \) is called the *proper time separation* of \( x_0 \) and \( x \).

The signature of the Lorentz inner product reverses some of the familiar inequalities from Euclidean geometry. The following results are Theorems 1.4.1 and Theorem 1.4.2 of [Nab4].

**Theorem 2.3.6.** (Reversed Schwartz Inequality) If \( v \) and \( w \) are timelike vectors in \( \mathcal{M} \), then

\[
\langle v, w \rangle^2 \geq \langle v, v \rangle \langle w, w \rangle
\]

and equality holds if and only if \( v \) and \( w \) are parallel.
Theorem 2.3.7. (Reversed Triangle Inequality) If \( v \) and \( w \) are timelike vectors with the same time orientation (that is, \( \langle v, w \rangle > 0 \)), then \( v + w \) is timelike and

\[
\tau(v + w) \geq \tau(v) + \tau(w).
\]

Equality holds if and only if \( v \) and \( w \) are parallel.

Theorem 2.3.7 extends to any finite sum of timelike vectors all of which have the same time orientation (Corollary 1.4.4 of [Nab4]).

Now suppose \( I \) is an interval in \( \mathbb{R} \) and \( \alpha : I \to M \) is a smooth curve in \( M \). Then \( \alpha \) is said to be spacelike, timelike, or null, respectively, if its velocity vector \( \alpha'(t) \), identified with a vector in \( M \), is spacelike, timelike or null, respectively, for each \( t \in I \).

Remark 2.3.6. Note that if \( I \) has endpoints, then “smooth” means that \( \alpha \) extends to an open interval on which it is \( C^\infty \).

A smooth timelike or null curve \( \alpha : I \to M \) in \( M \) is future directed (respectively, past directed) if \( \alpha'(t) \) is future directed (respectively, past directed) for every \( t \in I \). A future directed timelike curve is also called a timelike worldline, or the worldline of a material particle and is interpreted physically as the set of all events in the history of some material particle.

If \( \alpha : [a, b] \to M \) is a timelike worldline joining \( \alpha(a) \) and \( \alpha(b) \), then we define the proper time length \( L(\alpha) \) of \( \alpha \) by

\[
L(\alpha) = \int_a^b \sqrt{\langle \alpha'(t), \alpha'(t) \rangle} dt = \int_a^b \tau(\alpha'(t)) dt.
\]

The physical interpretation of \( L(\alpha) \) is a bit more subtle than one might expect and depends on additional physical input called the “Clock Hypothesis” (see pages 48-50 of [Nab4]). The generally accepted interpretation is that \( L(\alpha) \) is the time lapse between the events \( \alpha(a) \) and \( \alpha(b) \) as measured by an ideal standard clock carried along by the particle whose worldline is \( \alpha \). This is always less than or equal to the proper time separation of \( \alpha(a) \) and \( \alpha(b) \); the following result combines Theorem 1.4.6 and Theorem 1.4.8 of [Nab4].

Theorem 2.3.8. (Twin Paradox) Let \( \alpha : [a, b] \to M \) be a timelike worldline from \( \alpha(a) \) to \( \alpha(b) \). Then the displacement vector \( \alpha(b) - \alpha(a) \) is timelike and future directed and

\[
L(\alpha) \leq \tau(\alpha(b) - \alpha(a)).
\]

Equality holds if and only if \( \alpha \) is a parametrization of the straight line joining \( \alpha(a) \) and \( \alpha(b) \).
Exercise 2.3.5. Do a little outside reading and decide for yourself why we choose to call Theorem 2.3.8 the Twin Paradox. Also decide for yourself if there is anything paradoxical about it.

We now define, for any timelike worldline $\alpha : [a, b] \rightarrow M$, what is for most purposes its most convenient parametrization. The *proper time function* $\tau = \tau(t)$ on $[a, b]$ is defined by

$$\tau = \tau(t) = \int_{a}^{t} \tau(\alpha'(u)) du = \int_{a}^{t} \langle \alpha'(u), \alpha'(u) \rangle^{1/2} du$$

for every $t$ in $[a, b]$. Since $\alpha$ is timelike, $\frac{d\tau}{dt}$ is smooth and positive so the inverse of $\tau = \tau(t)$ exists and is smooth with a positive derivative. We can therefore parametrize $\alpha$ by $\tau$ and we will abuse the notation a bit and write this parametrization simply $\alpha(\tau)$. Physically, we are parametrizing the timelike worldline by time readings actually recorded along the worldline, assuming the time is set to zero at $\alpha(a)$. Relative to any admissible basis we write

$$\alpha(\tau) = x^\mu(\tau)e_\mu.$$

The velocity vector

$$\alpha'(\tau) = \frac{dx^\mu}{d\tau} e_\mu$$

is called the *4-velocity* of the timelike worldline and

$$\alpha''(\tau) = \frac{d^2x^\mu}{d\tau^2} e_\mu$$

is its *4-acceleration*.

Exercise 2.3.6. Prove each of the following.
1. $\langle \alpha'(\tau), \alpha'(\tau) \rangle = 1$ for all $\tau$ in $[0, L(\alpha)]$.
2. $\langle \alpha'(\tau), \alpha''(\tau) \rangle = 0$ for all $\tau$ in $[0, L(\alpha)]$.

Thus, the 4-velocity is a unit timelike vector and the 4-acceleration is spacelike at each point.

### 2.4 Lorentz and Poincaré Groups

In this section we will look a bit more closely at the groups that will be of particular interest to us. The proper, orthochronous Lorentz group $L_+^1$ was introduced in the
previous section and consists of all $4 \times 4$ real matrices $\Lambda = (\Lambda^\alpha{}_{\beta})_{\alpha,\beta=0,1,2,3}$ satisfying $\Lambda^T \eta \Lambda = \eta$, $\Lambda^0{}_{0} \geq 1$ and $\det \Lambda = 1$. Physically, these are the coordinate transformation matrices between two admissible frames of reference that agree on a common event as the spacetime origin. $\mathcal{L}^+$ inherits a topology as a subspace of the real general linear group $\text{GL}(4, \mathbb{R})$, which is an open subset of $\mathbb{R}^{4^2}$. With this topology $\mathcal{L}^+$ is a Hausdorff topological group. Being a closed subgroup of $\text{GL}(4, \mathbb{R})$, $\mathcal{L}^+$ is, in fact, a smooth submanifold and a Lie group (Theorem 1.1.2). In Section 2.6 we will show that $\mathcal{L}^+$ is diffeomorphic to $\mathbb{R}^3 \times \text{SO}(3)$.

In particular, $\mathcal{L}^+$ is 6-dimensional, connected, and has fundamental group $\mathbb{Z}_2$ (for the last statement see Appendix B of [Nab4]).

Admissible observers that do not share a common spacetime origin will assign coordinates that differ by a translation and a Lorentz transformation. Specifically, if $\mathcal{S}$ and $\hat{\mathcal{S}}$ are two such frames of reference assigning coordinates $x$ and $\hat{x}$, respectively, then there exists an $a \in \mathbb{R}^{1,3}$ and a $\Lambda \in \mathcal{L}^+$ such that $\hat{x} = a + \Lambda x$. The composition of two such transformations is given by

$$x \mapsto a_2 + A_2 x \mapsto a_1 + A_1 (a_2 + A_2 x) = (a_1 + A_1 a_2) + (A_1 A_2) x$$

and this is a transformation of the same type with $a = a_1 + A_1 a_2$ and $\Lambda = \Lambda_1 \Lambda_2$. Somewhat more formally we define, for each $(a, \Lambda) \in \mathbb{R}^{1,3} \times \mathcal{L}^+$ an affine mapping $(a, \Lambda) : \mathbb{R}^{1,3} \to \mathbb{R}^{1,3}$ by

$$(a, \Lambda)(x) = a + \Lambda x$$

for every $x \in \mathbb{R}^{1,3}$. Then

$$(a_1, \Lambda_1) \circ (a_2, \Lambda_2) = (a_1 + A_1 a_2, A_1 \Lambda_2).$$

This we recognize (Section 1.5) as the multiplication for the semi-direct product of $\mathbb{R}^{1,3}$ (regarded as an additive translation group) and $\mathcal{L}^+$ determined by the natural action of $\mathcal{L}^+$ on $\mathbb{R}^{1,3}$. We will suppress this natural action from the notation and write the semi-direct product simply as $\mathbb{R}^{1,3} \rtimes \mathcal{L}^+$. It is called the Poincaré group and is denoted

$$P^1 = \mathbb{R}^{1,3} \rtimes \mathcal{L}^+.$$  

$P^1$ has the topology and manifold structure of $\mathbb{R}^{1,3} \times \mathcal{L}^+$ and is a 10-dimensional, connected Lie group with fundamental group $\pi_1(P^1) \equiv \mathbb{Z}_2$. For the record we recall that the identity element in $P^1$ is $(0, \text{id}_{4 \times 4})$ and that the inverse of any $(a, \Lambda)$ in $P^1$ is given by

$$(a, \Lambda)^{-1} = (\Lambda^{-1}(-a), \Lambda^{-1}) = (-\Lambda^{-1}a, \Lambda^{-1}).$$
It will be convenient to have an explicit matrix model of $\mathcal{P}_+^1$ and this is easily done. Define a mapping $\Gamma : \mathcal{P}_+^1 \to \text{GL}(5, \mathbb{R})$ by

$$
\Gamma(a, \Lambda) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
a^0 & A^0_0 & A^0_1 & A^0_2 & A^0_3 \\
a^1 & A^1_0 & A^1_1 & A^1_2 & A^1_3 \\
a^2 & A^2_0 & A^2_1 & A^2_2 & A^3_3 \\
a^3 & A^3_0 & A^3_1 & A^3_2 & A^3_3
\end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \Lambda \end{pmatrix}
$$

Now identify $\mathbb{R}^{1,3}$ with the subspace of $\mathbb{R}^5$ consisting of all

$$
X = \begin{pmatrix}
x_0 \\
x^1 \\
x^2 \\
x^3
\end{pmatrix} = \begin{pmatrix} 1 \\ x \end{pmatrix}.
$$

Then

$$
\Gamma(a, \Lambda)X = \begin{pmatrix} 1 \\ 0 \\ \Lambda \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ a + \Lambda x \end{pmatrix}
$$

Exercise 2.4.1. Show that $\Gamma$ is a Lie group isomorphism of $\mathcal{P}_+^1$ onto its image in $\text{GL}(5, \mathbb{R})$.

General topological considerations imply that, because the fundamental groups of $\mathcal{L}_+^1$ and $\mathcal{P}_+^1$ are $\mathbb{Z}_2$, each of these groups has a “universal double covering group”. However, we will need explicit constructions of these so we will build them and explain as we go along what “universal double covering group” means (also see page 13). The construction depends on a rather remarkable reformulation of both Minkowski spacetime and the Lorentz group which we now describe.

We begin by considering the real vector space $H_2$ of $2 \times 2$ complex matrices $X$ that are Hermitian ($X^T = X$). This is precisely the set of matrices that can be written in the form

$$
X = \begin{pmatrix}
x_0 + x^3 & x^1 - ix^2 \\
x^1 + ix^2 & x_0 - x^3
\end{pmatrix} = x^0 \sigma_0 + x^1 \sigma_1 + x^2 \sigma_2 + x^3 \sigma_3 = x^a \sigma_a,
$$

where $x^0, x^1, x^2,$ and $x^3$ are real numbers, $\sigma_0$ is the $2 \times 2$ identity matrix and $\sigma_1, \sigma_2,$ and $\sigma_3$ are the Pauli spin matrices (see Exercise 1.2.6). Notice that

$$
\det X = \det \begin{pmatrix}
x_0 + x^3 & x^1 - ix^2 \\
x^1 + ix^2 & x_0 - x^3
\end{pmatrix} = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2.
$$

Define an inner product on $H_2$ by polarization of this quadratic form, that is,
\[ \langle X, Y \rangle_{H_2} = \frac{1}{4} [\det(X + Y) - \det(X - Y)] = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3. \]

Finally, notice that the map \( x = (x^0, x^1, x^2, x^3) \in \mathbb{R}^{1,3} \mapsto X =: x^0 \sigma_0 + x^1 \sigma_1 + x^2 \sigma_2 + x^3 \sigma_3 \in H_2 \), which is clearly linear, has an inverse given by

\[ x^\alpha = \frac{1}{2} \text{Trace} (\sigma_{\alpha} X), \quad \alpha = 0, 1, 2, 3, \]

We conclude that this map is an isomorphism of \( \mathbb{R}^{1,3} \) onto \( H_2 \) that sends the Lorentz inner product to the \( H_2 \)-inner product so that we are free to fully identify \( \mathbb{R}^{1,3} \) and \( H_2 \).

Next we consider the Lie group \( \text{SL}(2, \mathbb{C}) \) of \( 2 \times 2 \) complex matrices with determinant one. For every \( A \in \text{SL}(2, \mathbb{C}) \) we define a mapping \( M_A : H_2 \to H_2 \) by

\[ M_A(X) = AXA^T \]

for every \( X \in H_2 \).

**Exercise 2.4.2.** Show that \( M_A(X) \) is in \( H_2 \) and \( \det M_A(X) = \det X \) for every \( A \in \text{SL}(2, \mathbb{C}) \) and every \( X \in H_2 \).

But then \( M_A(X) \) can be uniquely written in the form

\[ M_A(X) = \begin{pmatrix} x^0 + \hat{x}^3 & \hat{x}^1 - i \hat{x}^2 \\ \hat{x}^1 + i \hat{x}^2 & x^0 - \hat{x}^3 \end{pmatrix} = \hat{x}^0 \sigma_0 + \hat{x}^1 \sigma_1 + \hat{x}^2 \sigma_2 + \hat{x}^3 \sigma_3 = \hat{x}^\alpha \sigma_\alpha, \]

where \( \hat{x}^0, \hat{x}^1, \hat{x}^2, \) and \( \hat{x}^3 \) are real numbers. Thus,

\[ (\hat{x}^0)^2 - (\hat{x}^1)^2 - (\hat{x}^2)^2 - (\hat{x}^3)^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \]

Consequently, the mapping \( x = (x^\alpha)_{\alpha = 0, 1, 2, 3} \mapsto \hat{x} = (\hat{x}^\alpha)_{\alpha = 0, 1, 2, 3} \) defined by

\[ \begin{pmatrix} \hat{x}^0 + \hat{x}^3 \\ \hat{x}^1 + i \hat{x}^2 \\ \hat{x}^0 - \hat{x}^3 \end{pmatrix} = A \begin{pmatrix} x^0 + x^3 \\ x^1 + x^2 \\ x^0 - x^3 \end{pmatrix} A^T, \]

which is clearly linear for each fixed \( A \in \text{SL}(2, \mathbb{C}) \), preserves the quadratic form \( \eta_{\alpha\beta} x^\alpha x^\beta \) and therefore, by polarization, preserves the Lorentz inner product \( \eta_{\alpha\beta} x^\alpha y^\beta \).

The mapping is therefore a general Lorentz transformation which we will denote \( \Lambda_A \). Notice that \( \Lambda_A^{-1} = \Lambda_A \) for every \( A \in \text{SL}(2, \mathbb{C}) \). One can write out the matrix \( \Lambda_A \) explicitly in terms of the entries in \( A \) and we will do so in a moment, but for many purposes it is more useful to note that

\[ (\Lambda_A)^\alpha_\beta = \frac{1}{2} \text{Trace} (\sigma_\alpha A \sigma_\beta A^T), \quad \alpha, \beta = 0, 1, 2, 3. \]

To see this we compute
\[
\frac{1}{2} \text{Trace} (\sigma_\alpha A\sigma_\beta \bar{A}^T) x^\beta = \frac{1}{2} \text{Trace} (\sigma_\alpha A(\lambda^\beta \sigma_\beta)\bar{A}^T) \\
= \frac{1}{2} \text{Trace} (\sigma_\alpha A\bar{X}A^T) \\
= \frac{1}{2} \text{Trace} (\sigma_\alpha M_A(X)) \\
= \xi^\alpha.
\]

Notice that \((A_A)^0_0 = \frac{1}{2} \text{Trace} (\bar{A}A^T)\) which is one-half the sum of the squared moduli of the entries in \(A\). This is positive so \(A_A\) is orthochronous. To see that \(A_A\) is proper we proceed as follows. In Exercise 2.6.3 you will show that \(\text{SL}(2, C)\) is diffeomorphic to \(R^3 \times SU(2)\).

\[
\text{SL}(2, C) \cong R^3 \times SU(2)
\]

Since \(SU(2)\) is homeomorphic to the 3-sphere \(S^3\) (Theorem 1.1.4 of [Nab2]) and \(S^3\) is connected and simply connected (pages 118-119 of [Nab2]), it follows that \(\text{SL}(2, C)\) is connected and simply connected (see Theorem 2.4.10 of [Nab2]). Now, being Lorentz transformations, every \(A_A\) has \(\det A_A = \pm 1\). But \(\det A_A\) is a continuous function of the entries in \(A\) so it must be constant on the connected space \(\text{SL}(2, C)\). When \(A\) is the \(2 \times 2\) identity matrix, \(A_A\) is the \(4 \times 4\) identity matrix and this has determinant 1. Consequently, \(\det A_A = 1\) for all \(A \in \text{SL}(2, C)\). We conclude that \(A_A \in L_+^1\) for every \(A \in \text{SL}(2, C)\). Thus, we have a mapping

\[
k : \text{SL}(2, C) \to L_+^1, \\
k(A) = A_A
\]

of \(\text{SL}(2, C)\) to \(L_+^1\).

Next we would like to show that \(k\) is a group homomorphism, that is,

\[
k(BA) = k(B)k(A)
\]

for all \(A, B \in \text{SL}(2, C)\). What we want to show then is that \(A_{BA} = A_B A_A\) and for this it will be enough to show that, for any \(x = (x^\gamma)_{\gamma=0,1,2,3} \in R^{1,3}\),

\[
(A_{BA})^\alpha_\beta x^\beta = (A_B)^\gamma_\alpha (A_A)^\alpha_\beta x^\beta.
\]

First note that, as we showed above,

\[
\bar{x}^\alpha = (A_A)^\alpha_\beta x^\beta = \frac{1}{2} \text{Trace} (\sigma_\alpha A\bar{X}A^T) = \frac{1}{2} \text{Trace} (\sigma_\alpha M_A(X))
\]

and similarly

\[
(A_B)^\gamma_\alpha \bar{x}^\alpha = \frac{1}{2} \text{Trace} (\sigma_\gamma M_B(\bar{X})) = \frac{1}{2} \text{Trace} (\sigma_\gamma (M_B \circ M_A)(X)).
\]
Finally,
\[
(A_B)^{\gamma \rho} \chi^{\beta} = \frac{1}{2} \text{Trace} \left( \sigma_{\gamma} (BA) X (BA)^T \right)
\]
\[
= \frac{1}{2} \text{Trace} \left( \sigma_{\gamma} B (AXA^{-T}) B^T \right)
\]
\[
= \frac{1}{2} \text{Trace} \left( \sigma_{\gamma} (M_B \circ M_A)(X) \right)
\]
\[
= (A_B)^{\gamma \rho} \chi^{\beta}
\]
\[
= (A_B)^{\gamma \rho} (A_A)^{\alpha \beta} \chi^{\alpha}
\]
as required.

Next we will show that the map \( \kappa \) is surjective and precisely two-to-one. For this it will be convenient to have in hand an explicit representation of \( A_A \) in terms of the entries
\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\]
of \( A \). Arriving at this is routine, albeit tedious. One can either write out the matrix product \(AXA^{-T}\) and read off the coefficients of the \( x^\gamma \) or use \((A_A)^{\gamma \rho} = \frac{1}{2} \text{Trace} (\sigma_{\gamma} A \sigma_{\rho} A^T)\). In either case the result is
\[
\begin{pmatrix} a^2 + |b|^2 + |c|^2 + |d|^2 & a c + a c + b \bar{d} + b \bar{d} + i(a \bar{c} + b \bar{d} - a \bar{c} - b \bar{d}) & |a|^2 + |b|^2 - |c|^2 - |d|^2 \\
\bar{c} b + a \bar{b} + \bar{c} d + c \bar{d} & a \bar{d} + \bar{a} d + b \bar{c} + b \bar{c} + i(\bar{a} d - \bar{b} c - a d + b c) & a \bar{b} + a \bar{b} - c \bar{d} - c \bar{d}
\end{pmatrix}
\]
\[
\begin{pmatrix} |a|^2 - |b|^2 + |c|^2 - |d|^2 & a \bar{c} + a \bar{c} - b \bar{d} - b \bar{d} + i(a \bar{c} + b \bar{d} - a \bar{c} - b \bar{d}) & |a|^2 - |b|^2 - |c|^2 + |d|^2
\end{pmatrix}
\]

We have already seen that \( \kappa(-A) = \kappa(A) \) and would now like to show that \( \kappa \) is precisely two-to-one. Since \( \kappa \) is a group homomorphism we need only show that the kernel of \( \kappa \) is \( \{ \pm I \} \), where \( I \) is the \( 2 \times 2 \) identity matrix.

Exercise 2.4.3. Equate the explicit matrix representation for \( A_A \) to the \( 4 \times 4 \) identity matrix and show that \( A = \pm I \).

Exercise 2.4.4. Let
\[
A_1(\theta) = \begin{pmatrix} \cosh(\theta) / 2 & \sinh(\theta) / 2 \\ \sinh(\theta) / 2 & \cosh(\theta) / 2 \end{pmatrix}
\]
Then \( A_1(\theta) \) is in \( \text{SL}(2, \mathbb{C}) \). Show that \( \kappa(A_1(\theta)) \) is the element of \( \mathcal{L}^+ \) representing a boost in the \( x^1 \)-direction.
\[
\kappa(A_1(\theta)) = A_1(\theta)
\]

If one has still not wearied of this laborious arithmetic one can check that, for any \( t \in [0, \pi] \) and any unit vector \( \mathbf{n} = (n^1, n^2, n^3) \) in \( \mathbb{R}^3 \), the matrix exponential
\[
\kappa(A_1(\theta)) = A_1(\theta)
\]
is in SU(2) \subseteq SL(2, \mathbb{C}) and its image under \( \kappa \) is the rotation in \( \mathcal{L}_+^1 \) given by

\[
R = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & R(n, t) \\
0 & 0 & 1
\end{pmatrix},
\]

where \( R(n, t) \) is the rotation in SO(3) described in Theorem A.2.2. Since every rotation in SO(3) is of this form we can conclude from this that \( \kappa \) maps the SU(2) subgroup of SL(2, \mathbb{C}) onto the rotation subgroup \( \mathcal{R} \) of \( \mathcal{L}_+^1 \). Then we appeal to Theorem 2.3.5 which asserts that any \( \Lambda \in \mathcal{L}_+^1 \) can be written as \( \Lambda = R_1 A_1(\theta)R_2 \), where \( R_1 \) and \( R_2 \) are rotations in \( \mathcal{L}_+^1 \) and to the fact that \( \kappa \) is a homomorphism to conclude that \( \kappa \) maps SL(2, \mathbb{C}) onto \( \mathcal{L}_+^1 \).

**Exercise 2.4.5.** Prove each of the following directly from the explicit representation for \( \Lambda_A \) in terms of the entries for \( A \).

\[
k(e^{i(-\frac{1}{2} \sigma_1)}) = \kappa\left( \begin{pmatrix}
\cos (t/2) & -i \sin (t/2) \\
-i \sin (t/2) & \cos (t/2)
\end{pmatrix}
\right) = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & \cos t \sin t \\
0 & \sin t & \cos t
\end{pmatrix}
\]

\[
k(e^{i(-\frac{1}{2} \sigma_2)}) = \kappa\left( \begin{pmatrix}
\cos (t/2) & -\sin (t/2) \\
\sin (t/2) & \cos (t/2)
\end{pmatrix}
\right) = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & \cos t \sin t \\
0 & \sin t & \cos t
\end{pmatrix}
\]

\[
k(e^{i(-\frac{1}{2} \sigma_3)}) = \kappa\left( \begin{pmatrix}
\cos (t/2) & 0 \\
0 & \cos (t/2)
\end{pmatrix}
\right) = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & \cos t \sin t \\
0 & \sin t & \cos t
\end{pmatrix}
\]

There is one final consequence we would like to draw from the fact that \( \kappa : SL(2, \mathbb{C}) \rightarrow \mathcal{L}_+^1 \) is a surjective homomorphism with discrete kernel \( \mathbb{Z}_2 = \{ \pm I \} \). From the explicit matrix expression for \( \kappa(A) \) it is clear that \( \kappa \) is continuous. By Corollary 1.1.3, it is smooth. Since SL(2, \mathbb{C}) and \( \mathcal{L}_+^1 \) are both connected we can apply Theorem 1.1.7 to conclude that \( \kappa \) is a smooth covering map. Since SL(2, \mathbb{C}) is simply connected, it is, in fact, the universal covering group of \( \mathcal{L}_+^1 \). Because \( \kappa \) is two-to-one, SL(2, \mathbb{C}) is called the universal double cover of \( \mathcal{L}_+^1 \). SU(2) is also simply connected so it is the universal double cover of the rotation group \( \mathcal{R} \equiv SO(3) \). The map \( \kappa \) itself is referred to either as the covering map or, on occasion, the spinor map.

All of this extends at once to the Poincaré group which, we recall, is the semi-direct product...
\[ \mathbb{P}^+_+ = \mathbb{R}^{1,3} \rtimes \mathcal{L}_+^1. \]

determined by the natural action of \( \mathcal{L}_+^1 \) on \( \mathbb{R}^{1,3} \). Now define inhomogeneous \( SL(2, \mathbb{C}) \), denoted ISL(2, \( \mathbb{C} \)), to be the semi-direct product
\[ \text{ISL}(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}), \]

where the action of \( SL(2, \mathbb{C}) \) on \( \mathbb{R}^{1,3} \) is determined by the covering map \( \kappa : SL(2, \mathbb{C}) \to \mathcal{L}_+^1 \) as follows.
\[ A \cdot a = \kappa(A)a = Aa \]
for all \( A \in SL(2, \mathbb{C}) \) and all \( a \in \mathbb{R}^{1,3} \).

**Exercise 2.4.6.** Show that the map \( \text{id}_{\mathbb{R}^{1,3}} \times \kappa : \text{ISL}(2, \mathbb{C}) \to \mathbb{P}^+_+ \) defined by
\[ (\text{id}_{\mathbb{R}^{1,3}} \times \kappa)(a, A) = (a, \kappa(A)) \]
for all \((a, A) \in \text{ISL}(2, \mathbb{C})\) is a smooth homomorphism of the semi-direct products with kernel equal to \( \pm \text{id}_{\text{ISL}(2, \mathbb{C})} \). Conclude that \( \text{ISL}(2, \mathbb{C}) \) is the universal double cover of \( \mathbb{P}^+_+ \).

**Exercise 2.4.7.** Show that the kernel of \( \text{id}_{\mathbb{R}^{1,3}} \times \kappa : \text{ISL}(2, \mathbb{C}) \to \mathbb{P}^+_+ \) is precisely the center of \( \text{ISL}(2, \mathbb{C}) \).

### 2.5 Poincaré Algebra

#### 2.5.1 Introduction

In Sections A.2 and A.3 we noted the intimate connection between the “infinitesimal symmetries” of a classical mechanical system and its conservation laws. To exploit similar ideas in the relativistic context we will need to understand the structure of the Lie algebras of the Lorentz and Poincaré groups. The proper, orthochronous Lorentz group \( \mathcal{L}_+^1 \) is given as a matrix group; it is the connected component containing the identity in the semi-orthogonal group \( O(1,3) \) (see Example 1.1.1 (6) and (7)) so these two have isomorphic Lie algebras. Once this Lie algebra is determined as a set of matrices by computing velocity vectors to smooth curves through the identity the Lie bracket is just the matrix commutator. The Poincaré group \( \mathbb{P}^+_+ \) is given as a semi-direct product, but we have described an explicit matrix model of it in Section 2.4 so we can follow the same procedure for it (a simpler example of this procedure is described in Example 1.1.2). Furthermore, the universal double cover of \( \mathcal{L}_+^1 \) is
2.5 Poincaré Algebra

SL(2, C) so these have isomorphic Lie algebras and, similarly, \( \mathcal{P}_+^1 \) and ISL(2, C) have isomorphic Lie algebras.

2.5.2 Lie Algebra of \( \mathcal{L}_+^1 \)

Exercise 2.5.1. The semi-orthogonal group O(1, 3) is the set of \( 4 \times 4 \) real matrices \( A \) satisfying \( A^T \eta A = \eta \) (Example 1.1.1 (6)). Follow the same procedure as in Example 1.1.2 to show that the Lie algebra of O(1, 3), and therefore of \( \mathcal{L}_+^1 \), is given by

\[
\mathfrak{o}(1, 3) = \{ X \in \mathfrak{gl}(4, \mathbb{R}) : X^T = -\eta X \eta \}
\]

and that these are precisely the real matrices of the form

\[
X = \begin{pmatrix}
0 & X^{01} & X^{02} & X^{03} \\
X^{01} & 0 & -X^{12} & X^{13} \\
X^{02} & X^{12} & 0 & -X^{23} \\
X^{03} & -X^{13} & X^{23} & 0
\end{pmatrix}.
\]

From the explicit description of the elements of the Lie algebra \( \mathfrak{o}(1, 3) \) in this Exercise we can read off a basis for \( \mathfrak{o}(1, 3) \).

\[
M_1 = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix},
M_2 = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0
\end{pmatrix},
M_3 = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
N_1 = \begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix},
N_2 = \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix},
N_3 = \begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}
\]

Notice that the \( M_j, j = 1, 2, 3 \), are skew-symmetric whereas the \( N_j, j = 1, 2, 3 \), are symmetric.

Computing the matrix commutators one finds that these basis elements satisfy the following commutation relations (\( \epsilon_{jkl} \) is the Levi-Civita symbol).

\[
[M_j, M_k] = \epsilon_{jkl}M_l, \quad j, k = 1, 2, 3 \tag{2.10}
\]

\[
[N_j, N_k] = -\epsilon_{jkl}M_l, \quad j, k = 1, 2, 3 \tag{2.11}
\]

\[
[M_j, N_k] = \epsilon_{jkl}N_l, \quad j, k = 1, 2, 3 \tag{2.12}
\]
Comparing the first of these with Exercise 1.1.2 we find that \( \{M_1, M_2, M_3\} \) generate a Lie algebra isomorphic to the Lie algebra \( \mathfrak{so}(3) \) of the rotation group. In particular, since the exponential map on \( \mathfrak{so}(3) \) is surjective (Theorem A.2.2) any rotation in \( \mathcal{L}_+^1 \) can be written in the form

\[
e^{\theta M_j} = e^{\theta M_1} + e^{\theta M_2} + e^{\theta M_3}
\]

for some real numbers \( \theta^j, j = 1, 2, 3 \). \( M_1, M_2, \) and \( M_3 \) are called the *generators of rotations* in \( \mathcal{L}_+^1 \).

**Exercise 2.5.2.** Show that

\[
e^{\theta M_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta \downarrow & -\sin \theta \downarrow \\ 0 & 0 & \sin \theta \downarrow & \cos \theta \downarrow \end{pmatrix}
\]

and then compute \( e^{\theta M_2} \) and \( e^{\theta M_3} \).

**Remark 2.5.1.** Except for the 0th row of zeros and 0th column of zeros, \( M_1, M_2, \) and \( M_3 \) are just the basis vectors \( \hat{X}_1, \hat{X}_2, \) and \( \hat{X}_3 \) for \( \mathfrak{so}(3) \) introduced in Exercise 1.1.2. These are, of course, technically different objects, but it is often convenient to simply identify \( M_j \) and \( \hat{X}_j \) for \( j = 1, 2, 3 \), so that \( e^{\theta M_j} \) is either a rotation in \( \mathcal{L}_+^1 \) or the same rotation in \( \text{SO}(3) \) depending on the context. We will even take this one step further and notice that these, in turn, can, by Exercise 1.2.6 (10), be identified with the basis vectors \( \hat{i}_2 \hat{j}_1 \) of \( \mathfrak{su}(2) \) in which case \( e^{\theta M_j} \) is an element of \( \text{SU}(2) \) which corresponds to a rotation via the double covering.

\[
M_1, M_2, M_3 \quad \leftrightarrow \quad X_1, X_2, X_3 \quad \leftrightarrow \quad \frac{i}{2} \hat{\sigma}_1, \frac{i}{2} \hat{\sigma}_2, \frac{i}{2} \hat{\sigma}_3
\]

**Exercise 2.5.3.** Show that

\[
e^{\ell N_1} = \begin{pmatrix} \cosh \ell \downarrow & \sinh \ell \downarrow & 0 & 0 \\ \sinh \ell \downarrow & \cosh \ell \downarrow & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

and then compute \( e^{\ell N_2} \) and \( e^{\ell N_3} \).

Motivated by Exercise 2.5.3, \( N_1, N_2 \) and \( N_3 \) are called the *generators of boosts* in \( \mathcal{L}_+^1 \). Notice, however, that \( N_1, N_2 \) and \( N_3 \) do not close under commutator since \( [N_j, N_k] = -\epsilon_{jkl} N_l \) and therefore, unlike \( M_1, M_2, \) and \( M_3 \), they do not span a Lie algebra. This is because the composition of two boosts in non-collinear directions
2.5 Poincaré Algebra

is not a boost, but is the composition of a boost and a rotation (called a Wigner rotation); there is an elementary discussion of this in [O’DV].

An arbitrary element of the Lie algebra of $L^+_+$ is of the form $\theta^j M_j + \xi^j N_j$ for real numbers $\theta^j, \xi^j, j = 1, 2, 3$. Consequently, every

$$e^{\theta^j M_j + \xi^j N_j}$$

(2.13)

is in $L^+_+$. In fact, it is true, conversely, that every element of $L^+_+$ can be written in the form (2.13). Stated otherwise, the exponential map on the Lie algebra of $L^+_+$ is surjective. This is not obvious, however, and we will simply refer to the proof available at http://www.cis.upenn.edu/cis610/cis61005sl8.pdf.

We now have an explicit description of the Lie algebra of $L^+_+$, but for various reasons that we will mention as we proceed, it is useful to obtain a few alternate descriptions of the generators of the Lorentz transformations. We begin by consolidating all of the matrices $M_j, N_j, j = 1, 2, 3$, into a single $4 \times 4$ skew-symmetric matrix $L_{\mu \nu}, \mu, \nu = 0, 1, 2, 3$, defined by

$$\begin{pmatrix}
0 & -N_1 & -N_2 & -N_3 \\
N_1 & 0 & -M_3 & M_2 \\
N_2 & M_3 & 0 & -M_1 \\
N_3 & -M_2 & M_1 & 0
\end{pmatrix}.$$

The entries of $(L_{\mu \nu})_{\mu, \nu=0,1,2,3}$ are themselves $4 \times 4$-matrices. Specifically,

$$N_j = L_{j0} = -L_{0j}, \quad j = 1, 2, 3$$

and

$$M_j = L_{jk} = -L_{kj}, \quad j = 1, 2, 3,$$

where $jkl$ is an even permutation of 123. For fixed $\mu$ and $\nu$ the entries in the matrix $L_{\mu \nu}$ will be designated $L_{\mu \nu}^{\alpha \beta}$. Thus, for example, $L_{23}^{\alpha \beta}$ are just the entries of $M_1$ so

$$L_{23}^{\alpha \beta} = \begin{cases} -1, & \text{if } \alpha = 2, \beta = 3 \\
1, & \text{if } \alpha = 3, \beta = 2 \\
0, & \text{otherwise.}
\end{cases}$$

Notice that this can be written

$$L_{23}^{\alpha \beta} = \eta_{3\alpha} \delta^\alpha_2 - \eta_{2\beta} \delta^\beta_3.$$

Exercise 2.5.4. Check a few more cases to persuade yourself that, for all $\mu, \nu = 0, 1, 2, 3$ and all $\alpha, \beta = 0, 1, 2, 3$,

$$L_{\mu \nu}^{\alpha \beta} = \eta_{\alpha \beta} \delta_\mu^\alpha - \eta_{\mu \nu} \delta_\nu^\alpha.$$
Then lower the index $\alpha$, that is, define $L_{\mu\nu\beta} = \eta_{\alpha\beta} L_{\mu\nu\gamma}$, and show that

$$L_{\mu\nu\beta} = \eta_{\mu\alpha} \eta_{\nu\beta} - \eta_{\mu\beta} \eta_{\nu\alpha}$$

for all $\mu, \nu, \alpha, \beta = 0, 1, 2, 3$.

We can now write all of the commutation relations (2.10), (2.11), and (2.12) as a single relation. For each fixed $\mu, \nu, \alpha, \beta = 0, 1, 2, 3$ we have

$$[L_{\mu\nu}, L_{\alpha\beta}] = -\eta_{\mu\alpha} L_{\nu\beta} + \eta_{\mu\beta} L_{\nu\alpha} + \eta_{\nu\alpha} L_{\mu\beta} - \eta_{\nu\beta} L_{\mu\alpha}.$$  \hspace{0.5cm} (2.14)

For example, if $\mu = 0$, $\nu = 1$, $\alpha = 0$, and $\beta = 2$,

$$[L_{01}, L_{02}] = [-N_1, -N_2] = [N_1, N_2] = -\epsilon_{123} M_3 = -M_3$$

and

$$-\eta_{00} L_{12} + \eta_{02} L_{10} + \eta_{10} L_{02} - \eta_{12} L_{00} = -\eta_{00} L_{12} = -L_{12} = -M_3$$

so (2.14) reduces to (2.11).

Exercise 2.5.5. Write out as many more of these as it takes to convince you that (2.14) contains all of the commutation relations (2.10), (2.11), and (2.12).

Exercise 2.5.6. Show that, if $X = (X^\mu)_{\mu,\nu=0,1,2,3}$ is in the Lie algebra of $\mathcal{L}_1^+$ and $\tilde{X}^{\mu\nu} = \eta^{\mu\nu} X^\alpha$, then the image $\Lambda$ of $X$ in $\mathcal{L}_1^+$ under the exponential map can be written

$$\Lambda = e^{\frac{1}{2} \tilde{X}^{\mu\nu} L_{\mu\nu}}.$$

There are advantages to viewing what we have just done from the complexified perspective (see Remark 1.1.2). For this we note that the complexification of the Lie algebra of $\mathcal{L}_1^+$ contains, in particular, the matrices

$$J_j = i M_j, \hspace{0.5cm} j = 1, 2, 3,$$

and

$$K_j = i N_j, \hspace{0.5cm} j = 1, 2, 3.$$

In terms of these the commutation relations (2.10), (2.11), and (2.12) take the form

$$[J_j, J_k] = i \epsilon_{jkl} J_l, \hspace{0.5cm} j, k = 1, 2, 3 \hspace{0.5cm} (2.15)$$

$$[K_j, K_k] = -i \epsilon_{jkl} J_l, \hspace{0.5cm} j, k = 1, 2, 3 \hspace{0.5cm} (2.16)$$
\[ [J_j, K_k] = i \epsilon_{jkl} K_l, \quad j, k = 1, 2, 3. \] (2.17)

**Note:** Unless some confusion is likely to arise we will adopt the usual custom of suppressing the subscript \( C \) in the notation \([, ]_C\) for the bracket of the complexification (Remark 1.1.2).

Notice that \( J_1 \hookrightarrow J_2 \hookrightarrow J_3 \hookrightarrow K_1 \hookrightarrow K_2 \) and \( K_3 \) can still be regarded as generators of the Lorentz transformations by simply including a factor of \(-i\) in the argument of the exponential map, that is, every element of \( \mathcal{L}_+^1 \) can be written in the form

\[ e^{-i \theta (J_j + i K_k)}. \] (2.18)

Next we consolidate \( J_1 \hookrightarrow J_2 \hookrightarrow J_3 \hookrightarrow K_1 \hookrightarrow K_2 \) and \( K_3 \) into a single skew-symmetric matrix of matrices given by

\[
(M_{\mu \nu})_{\mu, \nu = 0, 1, 2, 3} = \begin{pmatrix}
0 & -K_1 & -K_2 & -K_3 \\
K_1 & 0 & -J_3 & J_2 \\
K_2 & J_3 & 0 & -J_1 \\
K_3 & -J_2 & J_1 & 0
\end{pmatrix}
\]

and note that (2.14) becomes

\[ [M_{\mu \nu}, M_{\eta \beta}] = -i (\eta_{\mu \alpha} M_{\eta \beta} - \eta_{\eta \beta} M_{\mu \alpha} - \eta_{\nu \alpha} M_{\mu \beta} + \eta_{\nu \beta} M_{\mu \alpha}). \] (2.19)

If \( X = (X^\nu)_{\mu, \nu = 0, 1, 2, 3} \) is in the Lie algebra of \( \mathcal{L}_+^1 \) and \( X^\nu = \eta^{\nu \gamma} X^\gamma \), then the image \( \Lambda \) of \( X \) in \( \mathcal{L}_+^1 \) under the exponential map can be written

\[ \Lambda = e^{-i X^\nu M_{\nu \nu}}. \]

Another useful set of complex generators for \( \mathcal{L}_+^1 \) is obtained in the following way. Define

\[ S_j = \frac{1}{2} (J_j + i K_j), \quad j = 1, 2, 3 \]

and

\[ T_j = \frac{1}{2} (J_j - i K_j), \quad j = 1, 2, 3. \]

**Exercise 2.5.7.** Show that, in terms of these, the commutation relations (2.15), (2.16), and (2.17) decouple as follows.

\[ [S_j, S_k] = i \epsilon_{jkl} S_l, \quad j, k = 1, 2, 3 \] (2.20)
\[ [T_j, T_k] = i\epsilon_{jkl}T_l, \quad j, k = 1, 2, 3 \]  
(2.21)

\[ [S_j, T_k] = 0, \quad j, k = 1, 2, 3 \]  
(2.22)

### 2.5.3 Lie Algebra of \( \mathcal{P}_+ \)

We now turn to the Poincaré algebra. The Poincaré group \( \mathcal{P}_+ \) is given as a semi-direct product \( \mathbb{R}^{1,3} \rtimes \mathcal{L}_+^1 \) and one could appeal to general results on Lie algebras of semi-direct products (see pages 301-306 of [Nab5] for a brief description and an example or Section I.4 of [Knapp] for the details). We prefer to follow a more pedestrian route using the explicit matrix model of \( \mathcal{P}_+ \) constructed in Section 2.4.

Recall that \( \mathcal{P}_+ \) is isomorphic to the closed subgroup of \( GL(5 \rtimes \mathbb{R}) \) consisting of all

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\alpha^0 & A^0 & A^1 & A^2 & A^3 \\
\alpha^1 & A^0 & A^1 & A^2 & A^3 \\
\alpha^2 & A^0 & A^1 & A^2 & A^3 \\
\alpha^3 & A^0 & A^1 & A^2 & A^3
\end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a & \Lambda \end{pmatrix}
\]

where \( a \in \mathbb{R}^{1,3} \) and \( \Lambda \in \mathcal{L}_+^1 \). The Lie algebra is therefore a collection of real, \( 5 \times 5 \) matrices with the bracket given by matrix commutator and we would like to find generators and commutation relations for it. Taking \( a = 0 \) one obtains a closed subgroup of \( \mathcal{P}_+ \) isomorphic to \( \mathcal{L}_+^1 \) and we will abuse the notation a bit by continuing to denote this \( \mathcal{L}_+^1 \). Similarly, taking \( \Lambda = \text{id}_{4 \times 4} \) gives a closed subgroup isomorphic to \( \mathbb{R}^{1,3} \) which we also denote \( \mathbb{R}^{1,3} \). Since the underlying manifold of \( \mathcal{P}_+ \) is the product \( \mathbb{R}^{1,3} \times \mathcal{L}_+^1 \), the tangent space at the identity

\[
\begin{pmatrix}
1 & 0 \\
0 & \text{id}_{4 \times 4}
\end{pmatrix}
\]

is just the vector space direct sum of the tangent spaces to \( \mathbb{R}^{1,3} \) and \( \mathcal{L}_+^1 \) at the identity. We have already determined generators and commutation relations for the Lie algebra of the Lorentz group \( \mathcal{L}_+^1 \). Thought of as living in the Lie algebra of \( \mathcal{P}_+ \) these are

\[
\begin{pmatrix} 0 & 0 \\ 0 & M_j \
\end{pmatrix}
\]

and

\[
\begin{pmatrix} 0 & 0 \\ 0 & N_j \
\end{pmatrix}
\]
for \( j = 1, 2, 3 \) and we will persist in our abuse of the notation by writing these simply as \( M_j \) and \( N_j \), respectively. They satisfy the commutation relations (2.10), (2.11), and (2.12). The matrices

\[
O_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad O_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\]
\[
O_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad O_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}
\]

are generators for the Lie algebra of the translation subgroup \( \mathbb{R}^{1,3} \subseteq \mathfrak{p}_+^{4} \) and they satisfy the commutation relations

\[ [O_\mu, O_\nu] = 0, \quad \mu, \nu = 0, 1, 2, 3. \]

Inserting a factor of \( i \) into each of the generators \( M_j \) and \( N_j \), \( j = 1, 2, 3 \), we obtain the complex generators of rotations and boosts

\[ J_j = \begin{pmatrix} 0 & 0 \\ 0 & iM_j \end{pmatrix}, \quad j = 1, 2, 3 \]

and

\[ K_j = \begin{pmatrix} 0 & 0 \\ 0 & iN_j \end{pmatrix}, \quad j = 1, 2, 3 \]

and thereby the matrix \( (M_\mu)_{\mu,\nu=0,1,2,3} \) satisfying the commutation relations (2.19). For the \( O_\mu, \mu = 0, 1, 2, 3 \), we introduce a factor of \(-i\) and define

\[ P_\mu = -iO_\mu, \quad \mu = 0, 1, 2, 3. \quad (2.23) \]

These satisfy

\[ [P_\mu, P_\nu] = 0, \quad \mu, \nu = 0, 1, 2, 3. \]

All that remains is to compute the brackets \([M_\mu, P_\alpha]\) and we claim that these are given by

\[ [M_\mu, P_\alpha] = i(\eta_{\alpha\nu}P_\mu - \eta_{\mu\nu}P_\alpha), \quad \mu, \nu, \alpha = 0, 1, 2, 3. \]

Exercise 2.5.8. Check this for \( \mu = 1, \nu = 0, \alpha = 0 \) and as many other cases as your conscience requires.
We are now in a position to summarize the structure of the Poincaré algebra. This is a complex Lie algebra \( \mathfrak{p} \) of dimension 10 (the complexification of the Lie algebra of the Poincaré group \( \mathfrak{p}_\pm \)). It has complex generators \( P_\mu, \mu = 0, 1, 2, 3 \), and \( M_{\mu\nu} = -M_{\nu\mu} \), with \( \mu, \nu = 0, 1, 2, 3 \), and \( \mu \neq \nu \) determined by the following commutation relations.

\[
[P_\mu, P_\nu] = 0 \tag{2.24}
\]

\[
[M_{\mu\nu}, M_{\alpha\beta}] = -i (\eta_{\mu\alpha} M_{\nu\beta} - \eta_{\nu\beta} M_{\alpha\mu} - \eta_{\alpha\mu} M_{\nu\beta} + \eta_{\nu\beta} M_{\mu\alpha}) \tag{2.25}
\]

\[
[M_{\mu\nu}, P_\alpha] = i (\eta_{\alpha\mu} P_\nu - \eta_{\alpha\nu} P_\mu) \tag{2.26}
\]

The complex generators \( J_j \) and \( K_j \) of rotations and boosts, respectively, are given by

\[
J_j = \epsilon_{jkl} M_{kl}, \quad j = 1, 2, 3,
\]

and

\[
K_j = M_{j0}, \quad j = 1, 2, 3,
\]

so the commutation relations (2.24), (2.25), and (2.26) can be written in terms of these by making specific choices for the subscripts in \( M_{\mu\nu} \). For example,

\[
[J_1, J_2] = [M_{23}, M_{31}] = -i (\eta_{23} M_{31}) = -i (M_{12}) = i J_3.
\]

Continuing in this way one obtains the following set of commutation relations for the Poincaré algebra \( \mathfrak{p} \).

\[
[J_j, J_k] = i \epsilon_{jkl} J_l, \quad j, k = 1, 2, 3 \tag{2.27}
\]

\[
[J_j, K_k] = i \epsilon_{jkl} K_l, \quad j, k = 1, 2, 3 \tag{2.28}
\]

\[
[K_j, K_k] = -i \epsilon_{jkl} J_l, \quad j, k = 1, 2, 3 \tag{2.29}
\]

\[
[J_j, P_k] = i \epsilon_{jkl} P_l, \quad j, k = 1, 2, 3 \tag{2.30}
\]

\[
[K_j, P_k] = i P_0 \delta_{jk}, \quad j, k = 1, 2, 3 \tag{2.31}
\]

\[
[J_j, P_0] = [P_j, P_0] = [P_0, P_0] = 0, \quad j = 1, 2, 3 \tag{2.32}
\]
\begin{align}
[K_j, P_0] = iP_j, \quad j = 1, 2, 3 \tag{2.33}
\end{align}

In physics it is generally the commutation relations of \(p\) that play the most prominent role rather than any particular realization of them as operators. We have already seen one concrete representation of these relations as \(5 \times 5\) matrices, that is, as operators on \(\mathbb{R}^5\). We will conclude this section with another realization of \(p\), this time as operators on an infinite-dimensional vector space. We will find that these two manifestations of the Poincaré algebra will suggest a link between the abstract bracket structure of \(p\) and the physics of relativistic systems, both classical and quantum.

Before getting started, however, we should point out that in much of what follows we will often treat \(P_\mu\) and \(M_\mu^\nu\) as if they were the components of objects that lived in \(\mathbb{R}^{1,3}\) rather than as elements of the Lie algebra \(p\). We will, for example, “raise indices” with \(\rho\) to define \(P_\mu = \rho^\mu_\nu P_\nu\) and \(M_\mu^\nu = \eta^\rho_\sigma \eta^\sigma_\nu M_\rho^\mu\) and, in the next section, form products such as \(P_\mu P_\mu\). The motivation for this is as follows. We saw in Example 1.1.5 that the Poincaré group \(\mathbb{P}_+\) and, therefore its subgroup \(\mathbb{L}_+\), acts on \(\mathbb{P}_+\) by conjugation and that this induces a right action of \(\mathbb{L}_+\) on the Lie algebra \(p\). In particular, \(\Lambda \in \mathbb{L}_+ \subseteq \mathbb{P}_+\) acts on each matrix \(P_\nu^\mu\) by \(\Lambda^{-1} P_\nu^\mu A\). We will ask you to show now that this has the same effect as transforming the \(P_\nu^\mu\) as if they were the components of a 4-vector in \(\mathbb{R}^{1,3}\).

**Exercise 2.5.9.** Compute the indicated matrix products and show that
\[
\Lambda^{-1} P_\nu^\mu A = \Lambda_\nu^\rho P_\rho, \quad \mu = 0, 1, 2, 3. \tag{2.34}
\]

On the other hand, one thinks of (2.35) below as saying that the generators \(M_\nu^\mu\) transform as a second rank 4-tensor under \(\mathbb{L}_+\).

**Exercise 2.5.10.** Show that
\[
\Lambda^{-1} M_\nu^\mu A = \Lambda_\nu^\alpha \Lambda_\mu^\beta M_\rho^\alpha, \quad \mu, \nu = 0, 1, 2, 3. \tag{2.35}
\]

**Remark 2.5.2.** One often sees similar terminology used in the physics literature, but in a different context so we should explain. First recall (Section 1.1) that the action of a matrix Lie group \(G\) on its Lie algebra \(\mathfrak{g}\) by conjugation is called the adjoint action of \(G\) and is denoted \(Ad : G \rightarrow Aut(\mathfrak{g})\). The derivative of \(Ad\) at the identity \(e \in G\) determines an action \(ad = Ad_{e,c} : \mathfrak{g} \rightarrow Der(\mathfrak{g})\) of \(\mathfrak{g}\) on \(\mathfrak{g}\). The value of \(ad\) at \(X \in \mathfrak{g}\) is denoted \(ad_X : \mathfrak{g} \rightarrow \mathfrak{g}\) and is given by \(ad_X Y = [X,Y]\) for every \(Y \in \mathfrak{g}\). This is called the adjoint action of \(\mathfrak{g}\). Intuitively, \(\mathfrak{g}\) (the tangent space at \(e \in G\)) is thought of as an “infinitesimal” version of \(G\). An element of \(so(3)\), for example, is an “infinitesimal rotation”. From this point of view the adjoint action of \(\mathfrak{g}\) is an infinitesimal version of the adjoint action of \(G\).

Consider, for example, the adjoint action of \(\mathbb{L}_+ \subseteq \mathbb{P}_+\) on \(p\). Exercise 2.5.9 asserts that under the action \(Ad_{\Lambda+}\) of the Lorentz group on \(p\), the generators \(P_\nu^\mu\) transform
like a 4-vector so that, even though it lives in \( p \), one can treat \( P = (P^0, P^1, P^2, P^3) \) as
if it were a vector in \( \mathbb{R}^{1,3} \). Now think of the elements of the Lie algebra \( o(1, 3) \subseteq p \) of \( \mathcal{L}_1 \) as “infinitesimal Lorentz transformations”. The adjoint action of \( o(1, 3) \), given by bracket, is then thought of as the action of infinitesimal Lorentz transformations on \( p \) and one can ask how the elements of \( p \) transform under this action. For example, let’s consider the generators \( K_1, K_2, K_3 \) in \( p \) and ask how these transform under infinitesimal rotations. The generators of the rotations are \( J_1, J_2, \) and \( J_3 \) so we are interested in the action of \( J_j \) on \( K_k \), that is,

\[
\text{adj}_j K_k = [J_j, K_k].
\]

But, according to (2.28),

\[
\text{adj}_j K_k = i \epsilon_{jkl} K_l
\]

so \((K_1, K_2, K_3)\) transforms like a vector under infinitesimal rotations. Physicists are inclined to omit the word “infinitesimal” and to write

\[
K = (K_1, K_2, K_3)
\]
as a reminder that \( K \) behaves in some ways like a vector in \( \mathbb{R}^3 \).

**Exercise 2.5.11.** Check that the commutation relations (2.24), (2.25), and (2.26) are the same with all of the indices raised, that is,

\[
[P^\mu, P^\nu] = 0 \tag{2.36}
\]

\[
[M^{\mu\nu}, M^{\rho\beta}] = -i (\eta^{\mu\rho} M^{\nu\beta} - \eta^{\mu\beta} M^{\nu\rho} - \eta^{\nu\rho} M^{\mu\beta} + \eta^{\nu\beta} M^{\mu\rho}) \tag{2.37}
\]

\[
[M^{\mu\nu}, P^\rho] = i (\eta^{\nu\rho} P^\mu - \eta^{\mu\rho} P^\nu) \tag{2.38}
\]

**Example 2.5.1.** Now we move on to our second realization of \( p \). We will consider admissible coordinates \( x^0, x^1, x^2 \) and \( x^3 \) on \( \mathbb{R}^{1,3} \) and will let \( C^\infty(\mathbb{R}^{1,3}; \mathbb{C}) \) be the vector space of smooth, complex-valued functions on \( \mathbb{R}^{1,3} \). We will write \( \partial_\mu \) for \( \frac{\partial}{\partial x^\mu} \) and will raise the indices with \( \eta \) to obtain \( \partial^\mu = \eta^{\mu\alpha} \partial_\alpha \) so that \( \partial^0 = \partial_0 \) and \( \partial^i = -\partial_i, i = 1, 2, 3 \). Now define operators \( P^\mu \) and \( M^{\mu\nu} \), \( \mu, \nu = 0, 1, 2, 3 \), on \( C^\infty(\mathbb{R}^{1,3}; \mathbb{C}) \) as follows.

\[
P^\mu = i \partial^\mu, \quad \mu = 0, 1, 2, 3,
\]

and

\[
M^{\mu\nu} = x^\mu P^\nu - x^\nu P^\mu, \quad \mu, \nu = 0, 1, 2, 3.
\]
Thus,

$$P^0 = i\partial_0 \quad \text{and} \quad P^j = -i\partial_j, \quad j = 1, 2, 3$$

$$M^{jk} = i(x^k\partial_j - x^j\partial_k), \quad j, k = 1, 2, 3 \quad \text{and} \quad M^{0j} = -M^{j0} = -i(x^0\partial_j + x^j\partial_0), \quad j = 1, 2, 3.$$  

We claim that, when the bracket is taken to be the operator commutator, these operators satisfy the commutation relations defining the Poincaré algebra, that is, (2.36), (2.37), and (2.38). The first of these is clear from the equality of mixed second order partial derivatives. We will check just two of the remaining cases to see how things go and then leave the rest to you. Specifically, we will first verify (2.37) when \(\mu = 0, \nu = 1, \alpha = 0,\) and \(\beta = 2\). Notice that, in this case, the right-hand side of (2.37) evaluated at \(\varphi \in C^\infty(\mathbb{R}^{1,3}; \mathbb{C})\) reduces to

$$-i(\eta^{00}M^{12}\varphi) = x^2\partial_1\varphi - x^1\partial_2\varphi$$

since all of the remaining \(\eta\) factors are zero. To see that the left-hand side of (2.37) is the same we just compute as follows.

$$[M^{01}, M^{02}]\varphi = [-i(x^0\partial_1 + x^1\partial_0)][-i(x^0\partial_2 + x^2\partial_0)]\varphi$$

$$- [i(x^0\partial_2 + x^2\partial_0)][-i(x^0\partial_1 + x^1\partial_0)]\varphi$$

$$= -x^0x^0\partial_1\partial_2\varphi - x^0x^2\partial_1\partial_0\varphi - x^1x^0\partial_0\partial_2\varphi - x^1x^2\partial_1\partial_0\varphi$$

$$+ x^0x^0\partial_1\partial_0\varphi + x^0x^1\partial_2\partial_0\varphi + x^2x^0\partial_1\partial_0\varphi + x^2x^1\partial_0\partial_0\varphi$$

$$= x^2\partial_1\varphi - x^1\partial_2\varphi$$

as required.

Next we will check (2.38) when \(\mu = 1, \nu = 2,\) and \(\alpha = 1\). Evaluated at \(\varphi\), the right-hand side contains only one nonzero term, namely,

$$i(-\eta^{11}P^2\varphi) = i(-i\partial_2\varphi) = \partial_2\varphi.$$  

For the left-hand side we compute

$$[M^{12}, P^1]\varphi = M^{12}(P^1\varphi) - P^1(M^{12}\varphi)$$

$$= i(x^2\partial_1 - x^1\partial_2)(-i\partial_1\varphi) + i\partial_1(i(x^2\partial_1\varphi - x^1\partial_2\varphi))$$

$$= x^2\partial_1\partial_1\varphi - x^1\partial_2\partial_1\varphi - x^2\partial_1\partial_1\varphi + \partial_1(x^1\partial_2\varphi)$$

$$= \partial_2\varphi.$$  

(2.39)

*Exercise* 2.5.12. Complete the verification of (2.37) and (2.38).

In order to search for some underlying connection with physics in this realization we will compare it with what we know about classical and quantum mechanics.
from Appendix A. In the quantum arena, however, one should keep in mind that $C^\infty(\mathbb{R}^{1,3}; \mathbb{C})$ consists of smooth objects and therefore is not a Hilbert space so a more precise analogy will require self-adjoint extensions of the operators to something that is a Hilbert space. At the moment we are looking only for formal similarities that we can use as motivation later on. To facilitate the comparisons we will also adopt units in which $\hbar = 1$.

We begin by restricting our attention to some spatial cross section of $\mathbb{R}^{1,3}$ corresponding to a fixed time $x^0$ in a fixed admissible frame of reference. This is a copy of $\mathbb{R}^3$. The operators

$$P^j = -i \frac{\partial}{\partial x^j}, \quad j = 1, 2, 3,$$

on $C^\infty(\mathbb{R}^3; \mathbb{C})$ are then just the restrictions of the quantum mechanical momentum operators (Example A.4.1 and specifically (A.25)) on $L^2(\mathbb{R}^3)$ in the given frame of reference. This would seem to suggest that the generators $P^1, P^2$ and $P^3$ have something to do with linear momentum. The suggestion is strengthened when we recall that these are the generators of the spatial translations in the $\mathbb{R}^{1,3}$ subgroup of $\mathbb{P}_+^I$ and that, in classical mechanics, spatial translation symmetry implies conservation of linear momentum (Example A.2.1).

In the relativistic context, however, the three spatial components of non-relativistic momentum have no physical significance. Rather, they appear as the non-relativistic approximations to the spatial components of the momentum 4-vector whose time component is the total relativistic energy (see Remark 2.6.1). In classical mechanics conservation of total energy follows from time-translation symmetry (Example A.2.1) and $P^0$ is the generator of time translations in $\mathbb{R}^{1,3}$. Moreover, in non-relativistic quantum mechanics the total energy is given by the classical Hamiltonian operator $H$ on $L^2(\mathbb{R}^3)$ and this, according to the Schrödinger equation (see (A.16)), is related to the time evolution of the wave function by

$$i \frac{d}{dt}(\psi(t)) = H(\psi(t)),$$

where the time evolution $\psi(t)$ of the wave function is regarded as a curve in $L^2(\mathbb{R}^3)$ and the $t$-derivative is the tangent vector to this curve in $L^2(\mathbb{R}^3)$. One can regard the wave function as defined on $\mathbb{R}^{1,3}$ and, under certain circumstances, one can identify the $L^2(\mathbb{R}^3)$-derivative $\frac{d}{dt}(\psi(t))$ with the time partial derivative of $\psi$ in the given frame of reference (see Remark 6.2.14 of [Nab5] for more on this). The Schrödinger equation is then written

$$i \frac{\partial}{\partial \tau} \psi = H \psi$$

and this essentially identifies the operator

$$P^0 = i \frac{\partial}{\partial \tau}$$
with the total energy operator $H$. Now, the Schrödinger equation is *not* relativistically invariant so it has no real status on $\mathbb{R}^{1,3}$. However, the Relativity Principle asserts that all admissible frames are physically equivalent and this suggests that the Schrödinger equation should describe the non-relativistic limit of the time evolution in *any* admissible frame of reference. This, in turn, suggests that $P^0$ is the appropriate relativistic analogue of the 0th-component of the 4-momentum in quantum theory.

The preceding arguments were informal and, perhaps, not entirely persuasive, but we are simply trying to motivate “associating” the relativistic 4-momentum to the abstract generators $(P^0, P^1, P^2, P^3)$ of the Poincaré algebra $\mathfrak{p}$. Precisely how this rather vague “association” is made use of in practice will be addressed at the end of this section. In any case, we will proceed now to attempt something similar for the remaining generators $M^\mu_\nu, \mu, \nu = 0, 1, 2, 3$. Only six of these are independent so we will look just at the following generators.

$$M^{23} = i(x^3 \partial_2 - x^2 \partial_3), \quad M^{31} = i(x^1 \partial_3 - x^3 \partial_1), \quad M^{12} = i(x^2 \partial_1 - x^1 \partial_2)$$

and

$$M^{01} = -i(x^0 \partial_1 + x^1 \partial_0), \quad M^{02} = -i(x^0 \partial_2 + x^2 \partial_0), \quad M^{03} = -i(x^0 \partial_3 + x^3 \partial_0).$$

The appropriate interpretations for $M^{23}, M^{31},$ and $M^{12}$ seem clear since these are just the operators representing (orbital) angular momentum in quantum mechanics (Example A.4.1 with $\hbar = 1$). Except for a factor of $-i$ they also bear a striking resemblance to the infinitesimal generators (A.3), (A.4), and (A.5) of angular momentum in classical mechanics. This suggests that $(J^1, J^2, J^3) = (M^{23}, M^{31}, M^{12})$ should be “associated with” the components of angular momentum.

The operators describing $M^{01}, M^{02},$ and $M^{03}$ are less familiar. These correspond to the generators of boosts in $\mathfrak{p}^+_0$ and are the operators associated with what is called relativistic angular momentum. This is a topic that generally does not find its way into most introductions to special relativity (in particular, it will not be found in [Nab4]) and any discussion of it here would take us rather far afield. For those who are interested in pursuing this we can suggest Section 7.8 of [Gold] for the physicist’s perspective and Chapter VII of [Synge] for a geometrical treatment in much the same spirit as [Nab4]. We will take this rather subtle physics for granted and simply “associate” $(K^1, K^2, K^3) = (M^{01}, M^{02}, M^{03})$ with relativistic angular momentum.

**Exercise 2.5.13.** Regard

$$M^{23} = i(x^3 \partial_2 - x^2 \partial_3), \quad M^{31} = i(x^1 \partial_3 - x^3 \partial_1), \quad M^{12} = i(x^2 \partial_1 - x^1 \partial_2)$$

as operators on $L^2(\mathbb{R}^3)$ and let $\psi$ be a smooth element of $L^2(\mathbb{R}^3)$.

1. Prove the following commutation relations for these operators.
\[ [M^{23}, M^{31}] = i M^{12}, \quad [M^{31}, M^{12}] = i M^{23}, \quad [M^{12}, M^{23}] = i M^{31} \]

2. Suppose \( \psi \) depends only on distance to the origin, that is, \( \psi(x^1, x^2, x^3) = \psi(|x|) \). Show that \( M^{23}\psi = M^{31}\psi = M^{12}\psi = 0 \) so that these operators depend only on the angular coordinates in \( \mathbb{R}^3 \). This suggests rewriting the operators in spherical coordinates.

3. Introduce spherical coordinates \( \rho, \phi \) and \( \theta \) in \( \mathbb{R}^3 \) by
\[
    x^1 = \rho \sin \phi \cos \theta, \quad x^2 = \rho \sin \phi \sin \theta, \quad x^3 = \rho \cos \phi,
\]
where \( 0 \leq \phi < \pi \) and \( 0 \leq \theta < 2\pi \). Show that
\[
    M^{23} = i (\sin \theta \partial_\phi + \cot \phi \cos \theta \partial_\theta), \quad M^{31} = -i (\cos \theta \partial_\phi - \cot \phi \sin \theta \partial_\theta), \quad M^{12} = -i \partial_\theta.
\]
Notice that none of these depend on \( \rho \).

4. Define \( M^2 = (M^{23})^2 + (M^{31})^2 + (M^{12})^2 \) on the smooth elements of \( L^2(\mathbb{R}^3) \) and show that
\[
    M^2 = -\left( \frac{1}{\sin \phi} \partial_\phi (\sin \phi \partial_\phi) + \frac{1}{\sin^2 \phi} \partial^2_\theta \right).
\]

Note: Physicists generally write the operators \( M^{23}, M^{31}, M^{12} \) and \( M^2 \) as \( L_x, L_y, L_z \) and \( L^2 \), respectively.

5. Compare this with the usual expression for the Laplacian \( \Delta \) in spherical coordinates on \( \mathbb{R}^3 \) to show that
\[
    \Delta = \frac{1}{\rho^2} \partial_\rho (\rho^2 \partial_\rho) - \frac{1}{\rho^2} M^2
\]
so that \( M^2 \) is the “spherical part” of the Laplacian on \( \mathbb{R}^3 \).

6. Prove that
\[
    [M^2, M^{23}] = [M^2, M^{31}] = [M^2, M^{12}] = 0.
\]

Remark 2.5.3. All of the operators \( M^{23}, M^{31}, M^{12} \) and \( M^2 \) are essentially self-adjoint on the Schwartz space \( \mathcal{S}(\mathbb{R}^3) \) and so have unique self-adjoint extensions to \( L^2(\mathbb{R}^3) \). As usual, we will use the same symbols to denote the extensions. These correspond to quantum observables (the components of the orbital angular momentum and the squared magnitude of the total orbital angular momentum, respectively) so the spectrum of any one of these operators contains the set of possible measured values of the corresponding observable (Postulate QM2 of Appendix A.4). In particular, one is interested in the “eigenvalue problem” for each of these operators.
In Section 5.3 of [Nab5], one or the other, or both, of these will be discussed in some detail in essentially any basic quantum mechanics book, although perhaps not at a level of rigor that will satisfy a mathematician (see, for example, Chapter 14 of [Bohm] or Sections 3.5 and 3.6 of [Sak]). A rigorous proof of the result we need, but will simply quote can be found in Chapter II, Section 7, of [Prug].

Because \( M^2 \) commutes with each of \( M^2_1, M^2_3, \) and \( M^{12} \) (Exercise 2.5.13 (6)) one can actually find simultaneous eigenfunctions for \( M^2 \) and any one of the angular momentum components, but not more than one since the \( M^{ij} \) do not commute with each other (Exercise 2.5.13 (1)). Physicists interpret this to mean that one can “simultaneously measure” \( M^2 \) and any single one of the angular momentum components (see pages 248-252 of [Nab5] for more on the notion of simultaneous measurability, which is more subtle than one might expect). We will, in fact, describe an orthonormal basis for \( L^2(\mathbb{R}^3) \) consisting of eigenfunctions for both \( M^2 \) and \( M^{12} \). In particular, it will follow from this that the complete spectrum of each of these operators (possible measured values of the associated observables) consists precisely of the corresponding eigenvalues. To find such an orthonormal basis we begin by noting that, if the 2-sphere \( S^2 \) and the ray \((0,\infty)\) are given the measures \( \sin \phi \, d\phi \, d\theta \) and \( 4\pi \rho^2 \, d\rho \), respectively, then the product measure is just the Lebesgue measure on \( \mathbb{R}^3 \). Moreover, there is a unique unitary map \( \pi : L^2(S^2) \otimes L^2((0,\infty)) \to L^2(\mathbb{R}^3) \) of the Hilbert space tensor product \( L^2(S^2) \otimes L^2((0,\infty)) \) onto \( L^2(\mathbb{R}^3) \) satisfying \( \pi(f \otimes g)(\phi, \theta, \rho) = f(\phi, \theta) g(\rho) \) for all \( f \in L^2(S^2), \, g \in L^2((0,\infty)) \), \( (\phi, \theta) \in S^2 \) and \( \rho \in (0,\infty) \) (Chapter II, Theorem 6.9, of [Prug]). Consequently, if \( \{f_1, f_2, \ldots \} \) is an orthonormal basis for \( L^2(S^2) \) and \( \{g_1, g_2, \ldots \} \) is an orthonormal basis for \( L^2((0,\infty)) \), then \( \{f_j \otimes g_k : j, k = 1, 2, \ldots \} \) is an orthonormal basis for \( L^2(S^2) \otimes L^2((0,\infty)) \) (Chapter II, Theorem 6.10, of [Prug]).

We will briefly describe how to obtain such an orthonormal basis of simultaneous eigenfunctions for \( M^2 \) and \( M^{12} \). The eigenvalue problems we are interested in are

\[
\frac{1}{\sin \phi} \partial_\phi (\sin \phi \partial_\phi \psi) + \frac{1}{\sin^2 \phi} \partial^2_\theta \psi = -\lambda \psi \quad (2.40)
\]

and

\[
i \partial_\phi \psi = -\mu \psi \quad (2.41)
\]

and we will begin with (2.40). Separating variables \( \psi(\rho, \phi, \theta) = R(\rho)Y(\phi, \theta) \) this reduces to

\[
R(\rho) \left( \frac{1}{\sin \phi} \partial_\phi (\sin \phi \partial_\phi Y(\phi, \theta)) + \frac{1}{\sin^2 \phi} \partial^2_\theta Y(\phi, \theta) + \lambda Y(\phi, \theta) \right) = 0.
\]

Consequently, if one finds solutions to the eigenvalue problem
\[
\frac{1}{\sin \phi} \partial_\phi (\sin \phi \partial_\phi Y(\phi, \theta)) + \frac{1}{\sin^2 \phi} \partial_\theta^2 Y(\phi, \theta) + \lambda Y(\phi, \theta) = 0, \tag{2.42}
\]

then \( R(\rho) \) can be chosen arbitrarily. As it happens, (2.42) is an old and well-understood problem in partial differential equations and mathematical physics. To describe its solutions we must introduce some equally old and well-known functions. For each

\[ l = 0, 1, 2, \ldots \]

and each

\[ m = -l, -l + 1, \ldots, l, \]

we define a function \( Y_{lm}(\phi, \theta) \) as follows.

\[
Y_{lm}(\phi, \theta) = (-1)^m \left[ \frac{2l + 1}{4\pi} \frac{(l + |m|)!}{(l - |m|)!} \right]^{1/2} P_{lm}(\cos \phi) e^{im\theta}, \tag{2.43}
\]

where \( P_{lm}(u) \) are the associated Legendre functions of order \( m \) defined by the Rodrigues’ formula

\[
P_{lm}(u) = \frac{(1 - u^2)^{m/2}}{2^l l!} \frac{d^{l+m}}{du^{l+m}}(u^2 - 1)^l, \quad m = 0, 1, 2, \ldots.
\]

The functions \( Y_{lm}(\phi, \theta) \), which are clearly smooth on \( S^2 \), are called spherical harmonics and our interest in them arises from the following result (which is Theorem 7.1, Chapter II, of [Prug]).

**Theorem 2.5.1.** For each \( l = 0, 1, 2, \ldots \) and each \( m = -l, -l + 1, \ldots, l \), the spherical harmonic \( Y_{lm}(\phi, \theta) \) satisfies

\[
\frac{1}{\sin \phi} \partial_\phi (\sin \phi \partial_\phi Y_{lm}(\phi, \theta)) + \frac{1}{\sin^2 \phi} \partial_\theta^2 Y_{lm}(\phi, \theta) + l(l + 1)Y_{lm}(\phi, \theta) = 0. \tag{2.44}
\]

Furthermore, the functions \( \{ Y_{lm}(\phi, \theta) : l = 0, 1, 2, \ldots, m = -l, -l + 1, \ldots, l \} \) form an orthonormal basis for \( L^2(S^2) \).

Consequently, if \( \{ R_j(\rho) : j = 0, 1, 2, \ldots \} \) is any orthonormal basis for \( L^2((0, \infty)) \), then \( \{ R_j(\rho)Y_{lm}(\phi, \theta) : j = 0, 1, 2, \ldots, l = 0, 1, 2, \ldots, m = -l, -l + 1, \ldots, l \} \) is an orthonormal basis for \( L^2(\mathbb{R}^3) \) consisting of eigenfunctions of \( M^2 \).

\[
M^2(\{ R_j(\rho)Y_{lm}(\phi, \theta) \}) = l(l + 1)R_j(\rho)Y_{lm}(\phi, \theta)
\]

In particular, the eigenvalues of \( M^2 \) are

\[ l(l + 1), \quad l = 0, 1, 2, \ldots. \]
Exercise 2.5.14. Show that each $R_j(\rho)Y_{lm}(\phi, \theta)$ is also an eigenfunction of $M^{12}$ with eigenvalue $m$.

Finally, we should point out that had we not chosen to work in units for which $\hbar = 1$ the eigenvalues (possible measured values) of $M^2$ would have a factor of $\hbar^2$

$$l(l + 1)\hbar^2$$

and those of $M^{12}$ would have a factor of $\hbar$.

$$m\hbar$$

Now that we have gained some intuition for what the physical interpretation of the generators of the Poincaré algebra “should” be we need to draw some conclusions from it regarding quantum systems that admit a unitary representation of $\text{ISL}(2, \mathbb{C})$ and so are “relativistically invariant” in a sense determined by the representation (such systems are discussed in more detail in Section 2.7). Since it costs no more to do so we will describe the plan in more generality by considering an arbitrary matrix Lie group $G$ with Lie algebra $\mathfrak{g}$ and a unitary representation $\sigma : G \to \mathcal{U}(\mathcal{H})$ of $G$ on a complex, separable Hilbert space $\mathcal{H}$. Ideally, we would like to define a Lie algebra homomorphism of $\mathfrak{g}$ into a Lie algebra of self-adjoint operators on $\mathcal{H}$ so that we can identify the images of the generators with observables of a quantum system whose Hilbert space is $\mathcal{H}$ and then supply these with an appropriate physical interpretation. There are at least two obvious difficulties with such a plan. The first is that observables are generally unbounded operators so that all of the usual domain issues arise and the commutator of two observables need not be defined on a subspace of $\mathcal{H}$ larger than the trivial one. Moreover, even if one can get around these domain issues, the sad fact is that, even for bounded operators, the commutator of two self-adjoint operators is not self-adjoint. Fortunately, this last difficulty is rather easy to circumvent.

Exercise 2.5.15. Let $A$ and $B$ be bounded, symmetric operators on $\mathcal{H}$ so that $\langle A\psi, \phi \rangle = \langle \psi, A\phi \rangle$ and $\langle B\psi, \phi \rangle = \langle \psi, B\phi \rangle$ for all $\psi, \phi \in \mathcal{H}$. Show that the commutator $[A, B]$ is skew-symmetric, that is,

$$\langle [A, B]\psi, \phi \rangle = -\langle \psi, [A, B]\phi \rangle$$

for all $A, B \in \mathcal{H}$.

Skew-symmetric operators, however, are better behaved with respect to the formation of commutators.

Exercise 2.5.16. Let $C$ and $D$ be bounded, skew-symmetric operators on $\mathcal{H}$ so that $\langle C\psi, \phi \rangle = -\langle \psi, C\phi \rangle$ and $\langle D\psi, \phi \rangle = -\langle \psi, D\phi \rangle$ for all $\psi, \phi \in \mathcal{H}$. Show that $[C, D]$ is also skew-symmetric.
Moreover, there is a simple one-to-one correspondence between symmetric and skew-symmetric operators, even if they are unbounded.

Exercise 2.5.17. Let $A$ be a symmetric operator on $\mathcal{H}$ with domain $\mathcal{D}(A)$ so that $\langle A\psi, \phi \rangle = \langle \psi, A\phi \rangle$ for all $\psi, \phi \in \mathcal{D}(A)$. Show that $iA$ is skew-symmetric on $\mathcal{D}(A)$. Show, conversely, that if $A$ is skew-symmetric, then $-iA$ is symmetric.

All of this extends to (essentially) self-adjoint and (essentially) skew-adjoint operators on $\mathcal{H}$ and suggests that, for the purposes of introducing a commutator structure on observables, one should focus on the latter rather than the former. Then one need only cure the domain problems. This, however, can only be accomplished by an assumption about the operators of interest. We begin with a definition that addresses these issues.

Let $\mathcal{H}$ be a complex, separable Hilbert space. A collection $\mathcal{W}$ of operators on $\mathcal{H}$ is said to be a Lie algebra of operators on $\mathcal{H}$ if the following conditions are satisfied.

1. There exists a dense linear subspace $\mathcal{D}$ of $\mathcal{H}$ such that, for every $W \in \mathcal{W}$,
   a. $\mathcal{D} \subseteq \mathcal{D}(W)$,
   b. $W(\mathcal{D}) \subseteq \mathcal{D}$,
   c. $W$ is essentially skew-adjoint on $\mathcal{D}$.

2. If $W, W_1$ and $W_2$ are in $\mathcal{W}$ and $a \in \mathbb{R}$, then there exist operators $R, S, T \in \mathcal{W}$ such that, for every $\psi \in \mathcal{D}$,

   \[ W_1(\psi) + W_2(\psi) = R(\psi) \]

   \[ aW(\psi) = S(\psi) \]

   and

   \[ W_1(W_2\psi) - W_2(W_1\psi) = T(\psi). \]

We will write $R, S$ and $T$ as $W_1 + W_2, aW$ and $[W_1, W_2]$, respectively, with the understanding that they may be defined only on $\mathcal{D}$ and are assumed to be in $\mathcal{W}$.

Remark 2.5.4. Some caution is required since, for example, “$W_1 + W_2$” as it is being used here need not be the same as the sum of the operators $W_1$ and $W_2$ in that it may have a different domain. The terminology not withstanding, a “Lie algebra of operators” need not be a “Lie algebra”.

Every element of $\mathcal{W}$ has a unique extension to a skew-adjoint operator on $\mathcal{H}$ which we will denote by the same symbol. Multiplying any of these by $-i$ then gives a self-adjoint operator on $\mathcal{H}$.

What we would like now is an analogue for strongly continuous, unitary representations of a Lie group on an infinite-dimensional Hilbert space of the fact that
any representation of a matrix Lie group $G$ on a finite-dimensional Hilbert space gives rise to a representation of the corresponding Lie algebra $\mathfrak{g}$ simply by differentiation at the identity. More precisely, and more generally, one has the following well-known result (if it is not-so-well-known to you, see Theorem 3.18 of [Hall]).

Theorem 2.5.2. Let $G$ and $H$ be matrix Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, and suppose $\phi : G \to H$ is a Lie group homomorphism. Then there exists a unique real linear map $d\phi : \mathfrak{g} \to \mathfrak{h}$ such that

1. $d\phi([X, Y]) = [d\phi(X), d\phi(Y)]_\mathfrak{h}$ for all $X, Y \in \mathfrak{g}$.

2. $d\phi(X) = \frac{d}{dt} \phi(e^{tX}) \bigg|_{t=0} = \lim_{t \to 0} \frac{\phi(e^{tX}) - 1}{t}$ for all $X \in \mathfrak{g}$, and

3. $\phi(e^X) = e^{d\phi(X)}$ for every $X \in \mathfrak{g}$.

This result applies, in particular, to a representation $\sigma : G \to \text{GL}(V)$ of $G$ on a finite-dimensional Hilbert space $V$ to give a representation $d\sigma : \mathfrak{g} \to \text{gl}(V)$ of the Lie algebra of $G$. Segal has proved an analogue of this result for strongly continuous, unitary representations of $G$ on any complex, separable Hilbert space (Theorem 3.1 of [Segal2]), but we will not need to appeal to this. We will proceed toward the application we have in mind in the following way.

Let $\sigma : G \to \text{U}(\mathcal{H})$ be a strongly continuous, unitary representation of the matrix Lie group $G$ on the complex, separable Hilbert space $\mathcal{H}$. For each $X$ in the Lie algebra $\mathfrak{g}$ of $G$ the 1-parameter subgroup $t \mapsto e^{tX}$ of $G$ is mapped by $\sigma$ to a strongly continuous 1-parameter group $t \mapsto \sigma(e^{tX})$ of unitary operators on $\mathcal{H}$. According to Stone’s Theorem (Section VIII.4 of [RS1]) there exists a unique skew-adjoint operator $d\sigma(X)$ on $\mathcal{H}$ such that

$$\sigma(e^{iX}) = \exp (i d\sigma(X))$$

for each $i \in \mathbb{R}$. Here the exponential map $\exp$ is defined by the functional calculus (Theorem 5.5.8 of [Nab5]) in the following way. Since $t d\sigma(X)$ is skew-adjoint, $-i (t d\sigma(X))$ is self-adjoint and the functional calculus gives a unitary operator

$$\exp (i [-i (t d\sigma(X))]).$$

Since $i [-i (t d\sigma(X))] = t d\sigma(X)$ we take this to be the definition of $\exp (t d\sigma(X))$. $d\sigma(X)$ is given by

$$d\sigma(X) \psi = \frac{d}{dt} \sigma(e^{tX}) \psi \bigg|_{t=0} = \lim_{t \to 0} \frac{\sigma(e^{tX}) \psi - \psi}{t}$$

and its domain is the set of all $\psi \in \mathcal{H}$ for which this limit in $\mathcal{H}$ exists.

Next we must appeal to a rather deep theorem of Nelson [Nel11] which asserts that there exists a dense linear subspace $\mathcal{D}_\sigma$ of $\mathcal{H}$ with $\mathcal{D}_\sigma \subseteq D(d\sigma(X)) \forall X \in \mathfrak{g}$ that is invariant under each $d\sigma(X)$.
and on which each \( d\sigma(X) \) is essentially skew-adjoint. The route Nelson took toward this result is sketched on pages 292-295 of [Nab5] and an application to the Heisenberg algebra is described on pages 295-300 of [Nab5]. As usual we will use the same symbol \( d\sigma(X) \) for the unique skew-adjoint extension of \( d\sigma(X) \). With \( D \) we can define the Lie algebra of operators \( \mathcal{W}(D) \) on \( \mathcal{H} \) consisting of all skew-symmetric operators \( W \) on \( \mathcal{H} \) for which \( D \) is essentially skew-adjoint on \( \mathcal{D} \). The map \( d\sigma : g \to \mathcal{W}(D) \) then satisfies

\[
d\sigma(X + Y) \psi = d\sigma(X) \psi + d\sigma(Y) \psi \quad \forall X, Y \in g \quad \text{and} \quad \forall \psi \in \mathcal{D},
\]

\[
d\sigma(aX) \psi = ad\sigma(X) \psi \quad \forall X \in g \quad \forall a \in \mathbb{R} \quad \text{and} \quad \forall \psi \in \mathcal{D},
\]

and

\[
d\sigma([X, Y]) \psi = [d\sigma(X), d\sigma(Y)] \psi \quad \forall X, Y \in g \quad \text{and} \quad \forall \psi \in \mathcal{D}.
\]

d\sigma is called a realization of \( g \) by skew-adjoint operators on \( \mathcal{H} \). For each \( X \in g \), \(-id\sigma(X)\) is self-adjoint on \( \mathcal{H} \). We will describe a concrete example at the end of Section 2.8.

**Remark 2.5.5.** It is not uncommon to see \( d\sigma \) referred to as a “representation” of the Lie algebra \( g \). The terminology can be misleading since the set of operators \( d\sigma(X) \) need not form a Lie algebra.

Now suppose that \( \mathcal{H} \) is the Hilbert space associated to some quantum system and \( G \) is a symmetry group of that system represented by \( \sigma : G \to \mathcal{U}(\mathcal{H}) \) (see Remark A.4.2). Then, for each \( X \in g \), \(-id\sigma(X)\) is a self-adjoint operator on \( \mathcal{H} \) and therefore corresponds to some observable for the quantum system. If \( \{X_1, \ldots, X_n\} \) is a basis for \( g \) and if each of the self-adjoint operators \(-id\sigma(X_j), j = 1, \ldots, n, \) has been associated with some physical quantity, then we will say that the symmetry group \( G \) has been assigned a physical interpretation.

Return now to the case in which \( G = \mathcal{P}_+^1 \) is the Poincaré group (or its universal cover \( \text{ISL}(2, \mathbb{C}) \) which has the same Lie algebra \( \mathfrak{p} \)). The existence of a unitary representation \( \sigma : \mathcal{P}_+^1 \to \mathcal{U}(\mathcal{H}) \) expresses a form of relativistic invariance of the quantum system whose Hilbert space is \( \mathcal{H} \) (this is discussed in more detail in Section 2.7). The Lie algebra \( \mathfrak{p} \) has ten generators \( P^\mu \) and \( M^\mu_\nu = -M^\nu_\mu, \mu, \nu = 0, 1, 2, 3, \mu \neq \nu, \) which we have already associated with physical quantities. Specifically, in a fixed admissible frame of reference, \( P^1, P^2, \) and \( P^3 \) are associated with the linear momentum in the \( x^1, x^2 \) and \( x^3 \) directions, \( P^0 \) with the energy, \( M^{23}, M^{31} \) and \( M^{12} \) with the
components of angular momentum, and \( M^{01}, M^{02}, \) and \( M^{03} \) with the components of relativistic angular momentum. Recall that the \( P^\mu, \mu = 0, 1, 2, 3, \) transform under \( \mathcal{L}_+ \) as a 4-vector (Exercise 2.5.9) and the \( M^{\mu\nu} \) transform under \( \mathcal{L}_\pm \) as a 4-tensor (Exercise 2.35). Moreover, conjugating by elements of the Abelian translation group \( \mathbb{R}^{1,3} \) leaves both invariant. From this we conclude that the physical interpretations of the generators are the same in every admissible frame of reference. We are therefore led to postulate that the corresponding self-adjoint operators \( -\mathrm{id}(P^0), -\mathrm{id}(P^1), \ldots \) on \( \mathcal{H} \) represent the same physical quantities in the relativistically invariant quantum theory whose Hilbert space is \( \mathcal{H} \). Thus, for example, \( -\mathrm{id}(P^0) \) is the operator representing the total energy of the system, otherwise known as the Hamiltonian.

In the next section we will introduce the machinery necessary to extend these considerations to operators arising from “products” of generators of the form \( P_\mu P^\mu \) which do not live in the Lie algebra, but rather in what is called the “universal enveloping algebra”.

### 2.5.4 Universal Enveloping Algebra and Casimir Invariants

In this section we need to define what are called the “Casimir invariants” for the Poincaré algebra \( p \). These objects do not live in the Lie algebra. They are certain quadratic functions of the generators of \( p \) and a Lie algebra does not have enough structure to make sense of a quadratic function of its elements. The idea is to construct a certain unital, associative algebra \( \mathcal{U}(\mathfrak{p}) \) with its associated commutator Lie algebra structure (Example 1.1.4) and show that \( p \) can be embedded, as a Lie algebra, in \( \mathcal{U}(\mathfrak{p}) \). One can then use the ambient multiplicative structure of \( \mathcal{U}(\mathfrak{p}) \) to define the required quadratic functions. \( \mathcal{U}(\mathfrak{p}) \) is called the “universal enveloping algebra” of \( \mathfrak{p} \). Every Lie algebra \( \mathfrak{g} \) has one and we will begin with a brief discussion of how it is defined as a universal object, why it exists and how it can be described concretely. A good reference for all of the details is Chapter III of [Knapp].

Remark 2.5.6. Although much of what we have to say is true of both real and complex Lie algebras, we will generally want the field of scalars to be algebraically closed. For this reason we will consider only complex Lie algebras in this section. If the Lie algebra of interest at any particular moment happens to arise as the (real) Lie algebra of some Lie group we will therefore reserve the symbol \( \mathfrak{g} \) for the complexification of this real Lie algebra just as we wrote \( p \) for the complexification of the Lie algebra of the Poincaré group in the previous section.

We let \( \mathfrak{g} \) denote a finite-dimensional, complex Lie algebra. A universal enveloping algebra for \( \mathfrak{g} \) is a complex, unital, associative algebra \( \mathcal{U} \) (with its associated commutator Lie algebra structure) together with a Lie algebra homomorphism \( \iota : \mathfrak{g} \to \mathcal{U} \) with the following property. If \( \mathcal{A} \) is another complex, unital, associative algebra (with its associated commutator Lie algebra structure) and
\[ \phi : \mathfrak{g} \to \mathcal{A} \]

is another Lie algebra homomorphism, then there exists a unique algebra homomorphism

\[ \psi : \mathcal{U} \to \mathcal{A} \]

such that \( \psi(1_\mathcal{U}) = 1_\mathcal{A} \) and

\[ \phi = \psi \circ \iota. \]

It is not at all obvious, but the Poincaré-Birkhoff-Witt Theorem (Theorem 3.8 of [Knapp]) implies that \( \iota : \mathfrak{g} \to \mathcal{U} \) is injective so we can, and will, identify \( \mathfrak{g} \) as a Lie algebra with its image \( \iota(\mathfrak{g}) \) in \( \mathcal{U} \). In particular, the bracket on \( \mathfrak{g} \), however it was initially defined, can be viewed as the commutator bracket associated with the multiplication on \( \mathcal{U} \).

As usual, defining an object by a universal property does not guarantee that it exists, but it does guarantee that, if it exists, it must be unique. Thus, any two universal enveloping algebras for \( \mathfrak{g} \) must be isomorphic as unital, associative algebras and we are justified in denoting it \( \mathcal{U}(\mathfrak{g}) \) and calling it the universal enveloping algebra of \( \mathfrak{g} \). To settle the question of existence one must construct from \( \mathfrak{g} \) a unital, associative algebra and a Lie algebra homomorphism \( \iota \) from \( \mathfrak{g} \) into it that satisfies the universal property proposed in the definition. For this one begins with the tensor algebra

\[ \mathcal{T}(\mathfrak{g}) = \bigoplus_{k=0}^\infty \mathcal{T}^k(\mathfrak{g}) \]

of the vector space \( \mathfrak{g} \), where \( \mathcal{T}^0(\mathfrak{g}) = \mathbb{C}, \mathcal{T}^1(\mathfrak{g}) = \mathfrak{g} \) and, for \( k \geq 2, \)

\[ \mathcal{T}^k(\mathfrak{g}) = \mathfrak{g} \otimes \cdots \otimes \mathfrak{g} \]

is the \( k \)-th tensor power of \( \mathfrak{g} \) (there is a review of the tensor algebra in Appendix A of [Knapp]). This is a (graded) associative algebra with multiplication given by the tensor product \( \otimes \) and unit element \( 1 \in \mathcal{T}^0(\mathfrak{g}) = \mathbb{C} \). We let \( \mathcal{J}(\mathfrak{g}) \) denote the ideal in \( \mathcal{T}(\mathfrak{g}) \) generated by all elements of the form

\[ X \otimes Y - Y \otimes X - [X, Y] \]

for all \( X \) and \( Y \) in \( \mathfrak{g} = \mathcal{T}^1(\mathfrak{g}) \). The quotient \( \mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g}) \) is then a (graded) unital, associative algebra. Products in \( \mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g}) \) are generally written without the multiplication sign \( \otimes \) and the inclusion of \( \mathfrak{g} \) in \( \mathcal{T}(\mathfrak{g}) \) as a vector subspace induces a map denoted \( \iota : \mathfrak{g} \to \mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g}) \) that satisfies

\[ \iota([X, Y]) = \iota(X)\iota(Y) - \iota(Y)\iota(X) \]
for all $X,Y \in \mathfrak{g}$. The universal properties of the tensor algebra then imply that $\mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g})$ and $\iota : \mathfrak{g} \to \mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g})$ satisfy the defining properties of the universal enveloping algebra $\mathcal{U}(\mathfrak{g})$ (Proposition 3.3 of [Knapp]) and this establishes the existence of $\mathcal{U}(\mathfrak{g})$. In particular, $\iota$ is injective and therefore embeds $\mathfrak{g}$ in $\mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g})$ as a Lie subalgebra when $\mathcal{T}(\mathfrak{g})/\mathcal{J}(\mathfrak{g})$ is given its commutator Lie algebra structure.

Since $\iota$ embeds $\mathfrak{g}$ in $\mathcal{U}(\mathfrak{g})$ as a Lie algebra one generally suppresses all mention of it and simply regards $\mathfrak{g}$ as a Lie subalgebra of $\mathcal{U}(\mathfrak{g})$. Since $\mathfrak{g}$ then lives inside the associative algebra $\mathcal{U}(\mathfrak{g})$ we can form products of the elements of $\mathfrak{g}$ and these products will also live in $\mathcal{U}(\mathfrak{g})$. In particular, if $\{X_1,\ldots,X_n\}$ is a basis for $\mathfrak{g}$, then the monomials

$$X_{i_1}^j \cdots X_{i_k}^j$$

with $1 \leq k \leq n$, $j_l \geq 0$ for $1 \leq l \leq k$, and

$$i_1 < \cdots < i_k$$

are all in $\mathcal{U}(\mathfrak{g})$. According to Theorem 3.8 of [Knapp] these monomials actually form a basis for $\mathcal{U}(\mathfrak{g})$.

**Exercise 2.5.18.** Show that Lie algebra representations of $\mathfrak{g}$ are in one-to-one correspondence with algebra representations of $\mathcal{U}(\mathfrak{g})$. More precisely, prove each of the following.

1. Every Lie algebra representation $\phi : \mathfrak{g} \to \text{End}(V)$ of $\mathfrak{g}$ on a complex vector space $V$ extends to a unique algebra representation $\psi : \mathcal{U}(\mathfrak{g}) \to \text{End}(V)$ of $\mathcal{U}(\mathfrak{g})$ on $V$.

2. Every algebra representation $\psi : \mathcal{U}(\mathfrak{g}) \to \text{End}(V)$ of $\mathcal{U}(\mathfrak{g})$ on a complex vector space $V$ is the extension of a unique Lie algebra representation of $\mathfrak{g}$ on $V$.

This exercise applies, in particular, to the adjoint representation $ad$ of $\mathfrak{g}$ on $\mathfrak{g}$ which therefore extends to an algebra representation of $\mathcal{U}(\mathfrak{g})$ on $\mathfrak{g}$. We will denote this extension

$$ad : \mathcal{U}(\mathfrak{g}) \to \text{End}(\mathfrak{g})$$

and to call it the adjoint representation of $\mathcal{U}(\mathfrak{g})$.

**Remark 2.5.7.** Although we will not make use of it we mention that there is another, more analytic description of the universal enveloping algebra for the Lie algebra of a Lie group $G$. One can identify this Lie algebra with the left-invariant vector fields on $G$. Now, a vector field can be thought of as a derivation on the smooth functions defined on $G$, that is, as first order differential operator. Composing such operators generates left-invariant differential operators of higher order on $G$. Proposition 1.9, Chapter II, of [Helg] proves that the associative algebra generated by the left-invariant vector fields on $G$ and the identity map is isomorphic to $\mathcal{U}(\mathfrak{g})$. 

Now we will turn to the special case of the Poincaré algebra \( \mathfrak{p} \). A basis for \( \mathfrak{p} \) is given by

\[
X_1 = P_0, \quad X_2 = P_1, \quad X_3 = P_2, \quad X_4 = P_3,
\]
\[
X_5 = J_1 = M_{23}, \quad X_6 = J_2 = M_{31}, \quad X_7 = J_3 = M_{12},
\]
\[
X_8 = K_1 = M_{10}, \quad X_9 = K_2 = M_{20}, \quad X_{10} = K_3 = M_{30},
\]
so that the monomials (2.45) in these generators span \( \mathfrak{u}(\mathfrak{p}) \). We are particularly interested in two specific elements of \( \mathfrak{u}(\mathfrak{p}) \). The first is denoted \( P_2 \) and is defined as follows. For each \( \mu = 0, 1, 2, 3 \), define \( P_\mu \) by

\[
P_\mu = P_\mu P_\mu = P_0^2 - P_1^2 - P_2^2 - P_3^2.
\]

(2.47)

The second is denoted \( W^2 \) and is defined as follows. Begin by defining the components \( W_\mu \) of what is called the Pauli-Lubanski vector by

\[
W_\mu = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} M_\nu^\rho P_\sigma,
\]

(2.48)

where \( M_\nu^\rho = \eta^{\rho\gamma} \eta^{\nu\delta} M_{\delta\gamma} \) and \( \epsilon_{\mu\nu\rho\sigma} \) is the Levi-Civita symbol (1 if \( \mu\nu\rho\sigma \) is an even permutation of 0123, -1 if \( \mu\nu\rho\sigma \) is an odd permutation of 0123 and 0 otherwise). Raising the index we define

\[
W^\mu = \eta^{\mu\nu} W_\nu,
\]

so that \( W^0 = W_0 \) and \( W^j = -W_j, j = 1, 2, 3 \). Thus, for example,

\[
W^0 = W_0 = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} M_\nu^\rho P_\sigma = \frac{1}{2} \epsilon_{0123} M_{23}^1 P^3 + \frac{1}{2} \epsilon_{0213} M_{31}^2 P^3 + \frac{1}{2} \epsilon_{0321} M_{12}^3 P^3
\]
\[
+ \frac{1}{2} \epsilon_{0231} M_{23}^2 P^1 + \frac{1}{2} \epsilon_{0312} M_{31}^3 P^2 + \frac{1}{2} \epsilon_{0132} M_{12}^1 P^2
\]
\[
= \frac{1}{2} (1) M_{23}^1 P^3 + \frac{1}{2} (-1)(-M_{12}^1) P^3 + \frac{1}{2} (-1)(-M_{31}^2) P^3
\]
\[
+ \frac{1}{2} (1) M_{23}^2 P^1 + \frac{1}{2} (1) M_{31}^3 P^2 + \frac{1}{2} (-1)(-M_{12}^3) P^2
\]

and so

\[
W^0 = W_0 = M_{23}^1 P^3 + M_{31}^3 P^2 + M_{12}^1 P^2.
\]

Notice that

\[
W^\mu = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} M_\nu P_\sigma,
\]

where
\[ \epsilon^{\alpha\beta\gamma} = \eta^{\alpha\beta\mu} \eta^{\nu\gamma\delta} \epsilon_{\mu\nu\delta} = -\epsilon_{\alpha\beta\gamma}. \]

**Exercise 2.5.19.** Prove each of the following.

1. \( W_0 = W_0 = J_1 P_1 + J_2 P_2 + J_3 P_3 \)
2. \( W_1 = -W_1 = J_1 P_0 + K_2 P_3 - K_3 P_2 \)
3. \( W_2 = -W_2 = J_2 P_0 + K_3 P_1 - K_1 P_3 \)
4. \( W_3 = -W_3 = J_3 P_0 + K_1 P_2 - K_2 P_1 \)

**Exercise 2.5.20.** Prove that
\[ M^{\mu\nu} P_\mu P_\nu = 0. \] (2.49)

*Hint:* \([P_\mu, P_\nu] = 0 \) for all \( \mu, \nu = 0, 1, 2, 3 \).

Next we define
\[ W^2 = W_\mu W^\mu = W_0^2 - W_1^2 - W_2^2 - W_3^2. \] (2.50)

We will discuss the significance of \( P^2 \) and \( W^2 \) shortly, but first we will need a few useful relations (the brackets below refer to the commutator bracket in \( \mathcal{U}(\mathfrak{p}) \)).

\[ [W^{\mu}, P_\mu] = 0 \] (2.51)

\[ [W^{\mu}, P^\nu] = 0 \quad \forall \mu, \nu = 0, 1, 2, 3 \] (2.52)

\[ [W^{\mu}, M^{\alpha\beta}] = i (W^{\mu} \eta^{\alpha\beta} - W^{\beta} \eta^{\alpha\mu}) \quad \forall \mu, \nu, \alpha, \beta = 0, 1, 2, 3 \] (2.53)

\[ [W^{\mu}, W^{\nu}] = i \epsilon^{\mu\nu\rho\sigma} W_\rho P_\sigma \quad \forall \mu, \nu = 0, 1, 2, 3 \] (2.54)

For the proof of (2.51) we write
\[ W^{\mu} P_\mu = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} M^{\nu\rho} P_\sigma P_\mu = \frac{1}{2} M^{\nu\rho}(\epsilon^{\mu\nu\rho\sigma} P_\sigma P_\mu). \]

Now, for each fixed \( \nu \) and \( \rho \), \( \epsilon^{\mu\nu\rho\sigma} \) is skew-symmetric in \( \sigma \) and \( \mu \), whereas \( P_\sigma P_\mu \) is symmetric in \( \sigma \) and \( \mu \) (because \( [P_\sigma, P_\mu] = 0 \)). Consequently, the terms in the sum \( \epsilon^{\mu\nu\rho\sigma} P_\sigma P_\mu \) cancel in pairs for each \( \nu \) and \( \rho \) and (2.51) is proved.

To prove (2.52) we first note that, if \([ , , ]\) is the commutator bracket of any associative algebra \( \mathcal{A} \), then \([AB, C] = A[B, C] + [A, C]B \) for all \( A, B, C \in \mathcal{A} \) (just expand both sides). Now compute...
\[ [W^\mu, P^\nu] = \frac{1}{2} \epsilon^{\mu \nu \rho \sigma} [M_{\rho \sigma} P_\gamma, P_\nu] = \frac{1}{2} \epsilon^{\mu \nu \rho \sigma} M_{\rho \sigma} [P_\gamma, P_\nu] + \frac{1}{2} \epsilon^{\mu \nu \rho \sigma} [M_{\rho \sigma}, P_\nu] P_\gamma. \]

The first term is clearly zero since \([P_\gamma, P_\mu] = 0\) for all \(\gamma, \mu = 0, 1, 2, 3\). To see that the second term is also zero we first compute

\[ [M_{\rho \sigma}, P_\nu] = \eta^{\rho \nu} [M_{\rho \sigma}, P_\delta] = \eta^{\rho \nu} (\eta_{\delta \rho} P_\sigma - \eta_{\delta \sigma} P_\rho) = i (\delta^{\rho}_\sigma P_\rho - \delta^\sigma_\rho P_\rho). \]

Consequently,

\[ [W^\mu, P^\nu] = \frac{1}{2} i \epsilon^{\mu \nu \rho \sigma} \delta^{\rho}_\sigma P_\alpha P_\gamma - \frac{1}{2} i \epsilon^{\mu \nu \rho \sigma} \delta^\rho_\sigma P_\beta P_\gamma \]
\[ = \frac{1}{2} i \epsilon^{\mu \nu \rho \sigma} P_\alpha P_\gamma - \frac{1}{2} i \epsilon^{\mu \nu \rho \sigma} P_\beta P_\gamma \]
\[ = -i \epsilon^{\mu \nu \rho \sigma} P_\beta P_\gamma \]

which is zero because \(\epsilon^{\mu \nu \rho \sigma}\) is skew-symmetric in \(\beta\) and \(\gamma\), whereas \(P_\beta P_\gamma\) is symmetric in \(\beta\) and \(\gamma\). We will leave (2.53) and (2.54) for you.

**Exercise 2.5.21.** Prove each of the following.

1. \([W^\mu, M^{\rho \beta}] = i (W^\rho \eta^{\beta \sigma} - W^\sigma \eta^{\beta \rho})\)
2. \([W^\mu, W^\nu] = i \epsilon^{\mu \nu \rho \sigma} W_\rho P_\sigma\)

Although a bit more labor-intensive, similar arguments will establish

\[ W^2 = -\frac{1}{2} M_{\mu \nu} M^{\mu \nu} P^2 + M^{\mu \sigma} M_{\nu \rho} P_\mu P_\nu. \]  

(2.55)

The elements \(P^2\) and \(W^2\) of \(\mathfrak{u}(p)\) are called *Casimir invariants* of \(\mathfrak{p}\). The reason we care about them is that they commute with all of the generators of \(\mathfrak{p}\) in \(\mathfrak{u}(p)\), that is,

\[ [P_\mu, P^2] = 0 \quad \text{and} \quad [M_{\mu \nu}, P^2] = 0, \quad \mu, \nu = 0, 1, 2, 3 \]  

(2.56)

and

\[ [P_\mu, W^2] = 0 \quad \text{and} \quad [M_{\mu \nu}, W^2] = 0, \quad \mu, \nu = 0, 1, 2, 3 \]  

(2.57)

To prove (2.56) we compute

\[ [P_\mu, P^2] = [P_\mu, P_\rho P^\rho] = \eta^{\rho \nu} [P_\mu, P_\nu P_\gamma] = \eta^{\rho \nu} [P_\mu, P_\nu] P_\gamma + \eta^{\rho \nu} P_\rho [P_\mu, P_\gamma] = 0 \]

by (2.24). Next we compute, using (2.25) and (2.24),
\[ [M_{\mu\nu}, P^g] = [M_{\mu\nu}, P^g] = \eta^{\rho \gamma} [M_{\rho\mu}, P_\rho] P_\gamma + \eta^{\rho \gamma} P_\rho [M_{\rho\mu}, P_\gamma] \\
= \eta^{\rho \gamma} (i(\eta_{\mu\nu} P_\mu - \eta_{\mu\nu} P_\nu)) P_\gamma + \eta^{\rho \gamma} (i(\eta_{\mu\nu} P_\mu - \eta_{\mu\nu} P_\nu)) \\
= i (\delta^\gamma_{\mu\nu} P_\mu - \delta^\gamma_{\mu\nu} P_\nu + \delta^\gamma_{\mu\nu} P_\mu - \delta^\gamma_{\mu\nu} P_\nu) \\
= i ((P_\mu P_\nu - P_\nu P_\mu + P_\nu P_\mu - P_\mu P_\nu) \\
= 0. \]

**Exercise 2.5.22.** Prove (2.57). *Hint:* Use (2.55).

You are no doubt wondering about the significance of (2.56) and (2.57). What is implied by the fact that \( \mathbf{P} \) and \( \mathbf{W} \) commute in \( \mathfrak{u}(\mathfrak{p}) \) with all of the generators of \( \mathfrak{p} \)? Although we will not be in a position to offer a fully satisfactory explanation before Section 2.8 it may be helpful to consider a few finite-dimensional analogues of what we have done just to be aware of what we will try to mimic in the infinite-dimensional situation in which we find ourselves.

Let us suppose then that we have a matrix Lie group \( G \) with (complexified) Lie algebra \( \mathfrak{g} \) and a representation \( \sigma : G \rightarrow GL(\mathcal{V}) \) of \( G \) on some finite-dimensional complex vector space \( \mathcal{V} \). By Theorem 2.5.2, \( \sigma \) induces a representation \( d\sigma : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathcal{V}) \) of the Lie algebra on \( \mathcal{V} \). Now let \( \{X_1, \ldots, X_n\} \) be a basis for \( \mathfrak{g} \). Then \( \mathfrak{u}(\mathfrak{g}) \) is generated as an associative algebra by \( \{X_1, \ldots, X_n\} \) (see (2.45) and the remarks following it).

**Exercise 2.5.23.** Show that the Lie algebra representation \( d\sigma : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathcal{V}) \) extends to an algebra representation of \( \mathfrak{u}(\mathfrak{g}) \) on \( \mathcal{V} \), that is, to an algebra homomorphism of \( \mathfrak{u}(\mathfrak{g}) \) into the algebra \( \text{End}(\mathcal{V}) \) of endomorphisms of \( \mathcal{V} \).

We will use the same symbol

\[ d\sigma : \mathfrak{u}(\mathfrak{g}) \rightarrow \text{End}(\mathcal{V}) \]

for the algebra representation determined by \( d\sigma : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathcal{V}) \). Now suppose we have some element \( Z \) of \( \mathfrak{u}(\mathfrak{g}) \) that commutes with every generator \( X_\lambda \) in \( \mathfrak{u}(\mathfrak{g}) \). Since every element of \( \mathfrak{u}(\mathfrak{g}) \) is a linear combination of the monomials (2.45), \( Z \) must commute with everything in \( \mathfrak{u}(\mathfrak{g}) \), that is, \( Z \) is in the center \( \mathcal{Z}(\mathfrak{u}(\mathfrak{g})) \) of \( \mathfrak{u}(\mathfrak{g}) \). Since \( d\sigma \) is an algebra homomorphism, \( d\sigma(Z) \in \text{End}(\mathcal{V}) \) commutes with everything in the image of \( d\sigma \). If we now assume that \( \sigma : G \rightarrow GL(\mathcal{V}) \) is irreducible, so that \( d\sigma : \mathfrak{g} \rightarrow \mathfrak{gl}(\mathcal{V}) \) and also \( d\sigma : \mathfrak{u}(\mathfrak{g}) \rightarrow \text{End}(\mathcal{V}) \) are irreducible, we can appeal to the following version of Schur’s Lemma.

**Theorem 2.5.3.** *(Schur’s Lemma)* Let \( A \) be an associative algebra over \( \mathbb{C} \) and \( \mathcal{V} \) a finite-dimensional complex vector space. Suppose \( \rho : A \rightarrow \text{End}(\mathcal{V}) \) is an irreducible representation of \( A \) on \( \mathcal{V} \). If \( A \in \text{End}(\mathcal{V}) \) commutes with \( \rho(a) \in \text{End}(\mathcal{V}) \) for every \( a \in A \), then \( A = \lambda \text{id}_\mathcal{V} \) for some \( \lambda \in \mathbb{C} \).
space dual of we define the admissible basis present context the role of the configuration space is played by with points in \((T_xM)\) the tangent space of \(M\) classical mechanics, as we viewed it in Sections A.2 and A.3, begins with a configuration space \(M\), or by points in the cotangent bundle \(T^*M\), called the state space, or by points in the tangent bundle \(TM\), called the phase space. Elements of \(TM\) are pairs \((x, v_x)\), where \(x \in M\) and \(v_x\) is in the tangent space \(T_x(M)\) to \(M\) at \(x\), representing a velocity. Elements of \(T^*M\) are pairs \((x, \eta_x)\), where \(x \in M\) and \(\eta_x\) is in the dual \(T^*_x(M)\) of \(T_x(M)\), representing a conjugate momentum. When the configuration space is \(\mathbb{R}^n\) the velocities can be identified with points in \(\mathbb{R}^n\) and the conjugate momenta with points in its dual \((\mathbb{R}^n)^*\). In our present context the role of the configuration space is played by \(\mathcal{M}\). With a choice of admissible basis \(\{e_0, e_1, e_2, e_3\}\) this is identified with \(\mathbb{R}^{1,3}\). With this as motivation we define the momentum space \(\mathbb{P}^{1,3}\) associated with \(\{e_0, e_1, e_2, e_3\}\) to be the vector space dual of \(\mathbb{R}^{1,3}\).
The dual basis for $\mathbb{P}^{1,3}$ will be written $\{e^0, e^1, e^2, e^3\}$ and we will designate the points of $\mathbb{P}^{1,3}$ by $p = p_0 e^0$, or simply by their coordinates $(p_0, p_1, p_2, p_3) = (p_0, \mathbf{p})$. The natural $\mathcal{L}_+^1$-action on $\mathbb{R}^{1,3}$ gives rise to an $\mathcal{L}_+^1$-action on $\mathbb{P}^{1,3}$, called the \textit{contragredient action}, and defined as follows. If $p : \mathbb{R}^{1,3} \to \mathbb{R}$ is a linear functional in $\mathbb{P}^{1,3}$, then $\Lambda \cdot p$ is defined by

$$(\Lambda \cdot p)(x) = p(\Lambda^{-1} \cdot x)$$

for every $\Lambda \in \mathcal{L}_+^1$. In coordinates relative to $\{e^0, e^1, e^2, e^3\}$ this is given by

$$\Lambda \cdot p = (\Lambda^{-1})^T p.$$ Recalling from Exercise 2.3.3 that $(\Lambda^{-1})^T$ is in $\mathcal{L}_+^1$ whenever $\Lambda$ is in $\mathcal{L}_+^1$, we conclude that all of the following subsets of $\mathbb{P}^{1,3}$ are invariant under this $\mathcal{L}_+^1$-action. For any $m > 0$ define

$$X_m^+ = \{ p \in \mathbb{P}^{1,3} : p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2, \ p_0 > 0 \},$$

$$X_m^- = \{ p \in \mathbb{P}^{1,3} : p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2, \ p_0 < 0 \},$$

$$Y_m = \{ p \in \mathbb{P}^{1,3} : p_0^2 - p_1^2 - p_2^2 - p_3^2 = -m^2 \}$$

and, for $m = 0$, define

$$X_0^+ = \{ p \in \mathbb{P}^{1,3} : p_0^2 - p_1^2 - p_2^2 - p_3^2 = 0, \ p_0 > 0 \},$$

$$X_0^- = \{ p \in \mathbb{P}^{1,3} : p_0^2 - p_1^2 - p_2^2 - p_3^2 = 0, \ p_0 < 0 \},$$

$$X_0^0 = \{ 0 \}.$$ $X_m^+$ and $X_m^-$ for $m > 0$ are called the \textit{positive} and \textit{negative mass hyperboloids}, respectively. With the exception of a few remarks now and then we will concentrate almost exclusively on $X_m^+$.

\textbf{Remark 2.6.1.} We should say a word about the terminology. $X_m^+$ is called the \textit{positive mass hyperboloid} and the reason for the \textit{hyperboloid} designation is no doubt clear. However, the \textit{mass} $m$ is assumed positive for both $X_m^+$ and $X_m^-$ so \textit{positive} must refer to something else. To see what this might be we proceed as follows. Suppose $\alpha(\tau)$ is a timelike worldline parametrized by proper time and $m$ is a positive real number. Then the pair $(\alpha, m)$ is identified with a material particle whose worldline is $\alpha$ and whose mass is $m$. Since the 4-velocity $\alpha'(\tau)$ is a unit timelike vector at each point, $p = p(\tau) = m\alpha'(\tau)$ satisfies

$$\langle p, p \rangle = m^2.$$
In admissible coordinates this is just
\[ p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2. \]

Consequently, the 4-momentum of the particle lives in \( X_m = X^+_m \cup X^-_m \). Now we would like to rewrite \( p \) in terms of the instantaneous speed of the particle relative to the given frame of reference. We denote this
\[ \beta = \beta(x^0) = \sqrt{\left(\frac{dx^1}{dx^0}\right)^2 + \left(\frac{dx^2}{dx^0}\right)^2 + \left(\frac{dx^3}{dx^0}\right)^2} \]
and define \( \gamma = \gamma(x^0) = (1 - \beta^2(x^0))^{-1/2} \). We will omit the simple computations (which are available on pages 51-52 and 81-82 of [Nab4]) and just quote the result we need. Assuming \( p_0 > 0 \),
\[ p = (p_0, p_1, p_2, p_3) = m\gamma \left(1, \frac{dx^1}{dx^0}, \frac{dx^2}{dx^0}, \frac{dx^3}{dx^0}\right) = m\gamma (1, \mathbf{v}), \]
where \( \mathbf{v} \) is the ordinary velocity vector of the particle in the given frame. Now, since \( \beta(x^0) \) is in \((-1, 1)\) for each \( x^0 \) we can write a convergent binomial series expansion for \( \gamma \).
\[ \gamma = (1 - \beta^2)^{-1/2} = 1 + \frac{1}{2} \beta^2 + \frac{3}{8} \beta^4 + \cdots \]
If we write \( \mathbf{v} = (dx^1/dx^0, dx^2/dx^0, dx^3/dx^0) = (v^1, v^2, v^3) \) this gives
\[ p_i = mv^i + \frac{1}{2}mv^i \beta^2 + \frac{3}{8}mv^i \beta^4 + \cdots, \quad i = 1, 2, 3, \]
and
\[ p_0 = m + \frac{1}{2}m\beta^2 + \frac{3}{8}m\beta^4 + \cdots \]

The term \( mv^i \) in the expression for \( p_i \) is the \( i \)-th component of the Newtonian momentum in the given frame and the remaining terms are the relativistic corrections. On the other hand, the appearance of the term \( \frac{1}{2}m\beta^2 \) corresponding to the Newtonian kinetic energy in the expression for \( p_0 \) leads us to refer to \( p_0 \) as the total relativistic energy of the particle and denote it
\[ E = p_0 = m\gamma = m + \frac{1}{2}m\beta^2 + \frac{3}{8}m\beta^4 + \cdots \]
In particular, \( X^+_m \) is the surface in \( \mathbb{P}^{1,3} \) containing the 4-momenta of particles with mass \( m > 0 \) and positive energy. This would lead one to identify \( X^-_m \) with the surface in \( \mathbb{P}^{1,3} \) containing the 4-momenta of particles with negative energy. Classically one would simply ignore \( X^-_m \) as being unphysical. In quantum theory the situation
2.6 Momentum Space

is not so simple, but the physical interpretation of particles with negative energy involves subtleties that we will have no need to get into here and this accounts for our particular interest in \( X^+_m \).

We would like to point out one other consequence of the identification of \( E \) and \( p_0 \). We will call \( p = (p_1, p_2, p_3) \) the \textit{relativistic 3-momentum} of the particle in the given frame of reference and will write \( \| p \|^2 \) for \( p_1^2 + p_2^2 + p_3^2 \). Then

\[
E^2 = \| p \|^2 + m^2. \tag{2.58}
\]

This is called the \textit{Einstein energy-momentum relation}. When one uses traditional units of time rather than \( x^0 = ct \) (2.58) becomes

\[
E^2 = \| p \|^2 c^2 + m^2 c^4. \tag{2.59}
\]

It is worth pointing out that when \( \beta = 0 \) and \( E = p_0 > 0 \) this reduces to

\[
E = mc^2
\]

which it is possible you have seen before.

In Section 2.8 we will require some fairly detailed information about \( \mathbb{P}^{1,3} \) so we will devote the remainder of this section to deriving what we need. We have already seen that the positive mass hyperboloid \( X^+_m \) in \( \mathbb{P}^{1,3} \) is invariant under the \( \mathcal{L}_+^1 \)-action on \( \mathbb{P}^{1,3} \). Now we will show that it is the complete orbit of any one of its points under this action. In other words, \( \mathcal{L}_+^1 \) acts transitively on \( X^+_m \).

Let \( p = (p_0, p_1, p_2, p_3) \) be an arbitrary point in \( X^+_m \) \((m > 0)\). Since \( X^+_m \) is invariant under \( \mathcal{L}_+^1 \), any element of the \( \mathcal{L}_+^1 \)-orbit of \( p \) is in \( X^+_m \). We will show, conversely, that any element \( q = (q_0, q_1, q_2, q_3) \) of \( X^+_m \) is in the orbit of \( p \). Notice that it is enough to find a \( \Lambda \in \mathcal{L}_+^1 \) with \( \Lambda p = q \) since then \( (\Lambda^{-1})^T \in \mathcal{L}_+^1 \) and \( (\Lambda^{-1})^T \cdot p = q \). We first show that there is an element of \( \mathcal{L}_+^1 \) that sends \((m, 0, 0, 0)\) to \( p \). We will apply Exercise 2.3.4. Take \( \mu = (p_0^2 - m^2)^{1/2}/m \) and \( \nu = p_0/m \). Notice that \( p_0^2 - m^2 = p_1^2 + p_2^2 + p_3^2 = \| p \|^2 \) so \( \mu \) is real. Moreover, \( \nu > 0 \) since \( p \in X^+_m \) and \( \nu^2 - \mu^2 = 1 \). With these choices \( \Lambda_{\mu, \nu} \) sends \((m, 0, 0, 0)\) to \((p_0, \| p \|, 0, 0)\). Since the rotation group \( \text{SO}(3) \) acts transitively on the sphere of radius \( \| p \| \) in \( \mathbb{R}^3 \) (see pages 90-91 of [Nab2]) there is a rotation \( R \in \mathbb{R} \) that carries \((p_0, \| p \|, 0, 0)\) to \((p_0, p_1, p_2, p_3)\). Thus, \( R\Lambda_{\mu, \nu} \in \mathcal{L}_+^1 \) sends \((m, 0, 0, 0)\) to \((p_0, p_1, p_2, p_3)\), as required. Since \( p \) was arbitrary we have shown that, for any \( p, q \in X^+_m \), there exist \( \Lambda_p, \Lambda_q \in \mathcal{L}_+^1 \) with \( \Lambda_p(m, 0, 0, 0) = p \) and \( \Lambda_q(m, 0, 0, 0) = q \). Consequently, \( \Lambda = \Lambda_q\Lambda_p^{-1} \) carries \( p \) to \( q \). Every \( q \) in \( X^+_m \) is in the \( \mathcal{L}_+^1 \)-orbit of \( p \), as required. Analogous arguments starting with \((-m, 0, 0, 0)\) show that \( X^-_m \) is the complete orbit of any one of its points.

\textit{Exercise 2.6.1.} Prove the same result for \( X^+_0 \) by taking \( \mu = (p_0^2 - 1)/2p_0 \) and \( \nu = (p_0^2 + 1)/2p_0 \) and looking at the image of \((1, 1, 0, 0)\) under \( \Lambda_{\mu, \nu} \). Find an analogous argument for \( X^-_0 \).
Exercise 2.6.2. Prove the same result for \( Y_m \) by taking \( \mu = p_0/m \) and \( \nu = (p_0^2 + m^2)/m \) and looking at the image of \((0, m, 0, 0)\) under \( A_{\mu, \nu} \).

Since every point of \( \mathbb{P}^{1,3} \) is exactly one of \( X^\pm_m, Y_m, X^\pm_0 \) or \( X^0_0 \) these are, in fact, all of the orbits of \( \mathcal{L} \) in \( \mathbb{P}^{1,3} \).

Next we will need to determine the isotropy subgroup \( H_{P_0} \) of \( P_0 = (m, 0, 0, 0) \in X^+_m \) under the \( \mathcal{L} \)-action. Thus, we are looking for all of the \( \lambda \) for which \( \lambda \cdot P_0 = P_0 \), that is, for which \( (\lambda^{-1})^T P_0 = P_0 \). If \( \lambda = (A^\alpha)_{\alpha, \beta = 0, 1, 2, 3} \), then

\[
(\lambda^{-1})^T = \eta \Lambda \eta = \begin{pmatrix}
A^0_0 & -A^0_1 & -A^0_2 & -A^0_3 \\
-A^1_0 & A^1_1 & A^1_2 & A^1_3 \\
-A^2_0 & A^2_1 & A^2_2 & A^2_3 \\
-A^3_0 & A^3_1 & A^3_2 & A^3_3
\end{pmatrix}
\]

Applying this to \((m, 0, 0, 0)\) gives

\[
\begin{pmatrix}
A^0_0 & m \\
-A^1_0 & 0 \\
-A^2_0 & 0 \\
-A^3_0 & 0
\end{pmatrix}
\]

which is equal to

\[
\begin{pmatrix}
m \\
0 \\
0 \\
0
\end{pmatrix}
\]

if and only if the first column of \((\lambda^{-1})^T\) is

\[
\begin{pmatrix}1 \\
0 \\
0 \\
0
\end{pmatrix}
\]

and this is the case if and only if \((\lambda^{-1})^T\) is in the rotation subgroup \( \mathcal{R} \) of \( \mathcal{L} \) (Theorem 2.3.4). Consequently,

\[ H_{P_0} = \mathcal{R} \cong SO(3). \]

The same argument shows that, for \( m > 0 \), the isotropy subgroup of \((-m, 0, 0, 0)\) in \( X^+_m \) is \( \mathcal{R} \).

Remark 2.6.2. The isotropy subgroups for \( Y_m \) \((m > 0)\) can be obtained in a similar fashion, while those of \( X^+_0 \) are somewhat more involved. Since we will require only the \( X^+_m \) case we will simply refer those interested in seeing the results for \( Y_m \) and \( X^+_0 \) to pages 340-341 of [Vara].
2.6 Momentum Space

We have shown that \( L^+ \) acts transitively on the left on \( X^+ \) with isotropy subgroup \( R \) at \((m, 0, 0, 0)\). Consequently, the homogeneous manifold \( L^+ / R \) is diffeomorphic to \( X^+ \) (Theorem 1.1.6) which, in turn, is diffeomorphic to \( R^3 \) via the projection

\[
\pi_+ : X^+ \rightarrow R^3 \\
\pi_+(p_0, p_1, p_2, p_3) = (p_1, p_2, p_3).
\]

Since \( R^3 \) is connected and \( SO(3) \) is connected, it follows that \( L^+ \) is also connected (Proposition 1.6.5 of [Nab2]). We can say more, however. In Section 1.4 we showed that the natural projection \( \pi : L^+_+ \rightarrow L^+_+ / R \) has the structure of a principal \( R \)-bundle over \( L^+_+ / R \). But \( L^+_+ / R \simeq R^3 \) and every principal bundle over any \( R^n \) is trivial so

\[
L^+_+ \simeq R^3 \times SO(3).
\]

From this we conclude that, not only is \( L^+_+ \) connected, but that its fundamental group is \( Z_2 \) (Theorem 2.4.5, Theorem 2.4.10, and Appendix A of [Nab2]).

Let \( \kappa : SL(2, \mathbb{C}) \rightarrow L^+_+ \) be the double covering map described in Section 2.4. Then \( SL(2, \mathbb{C}) \) acts on \( R^{1,3} \) via the \( L^+_+ \)-action, that is, by

\[
A \cdot p = \kappa(A) \cdot p = (\kappa(A)^{-1})^T p
\]

for every \( A \in SL(2, \mathbb{C}) \) and every \( p \in R^{1,3} \). Since \( \kappa \) is surjective the orbits of the \( SL(2, \mathbb{C}) \)-action are the same as those of the \( L^+_+ \)-action and \( SL(2, \mathbb{C}) \) acts transitively on \( X^+ \). The isotropy group of any point in \( X^+ \) is the inverse image of the rotation group \( R \) in \( L^+_+ \) under \( \kappa \) and, as we have seen in Section 2.4, this is isomorphic to \( SU(2) \). Thus, just as for \( L^+_+ / R \), the homogeneous manifold \( SL(2, \mathbb{C}) / SU(2) \) is diffeomorphic to \( X^+ \) which, in turn, is diffeomorphic to \( R^3 \).

**Exercise 2.6.3.** Show that \( SL(2, \mathbb{C}) \) is diffeomorphic to \( R^3 \times SU(2) \).

According to Theorem 1.1.6 one can describe a diffeomorphism of \( SL(2, \mathbb{C}) / SU(2) \) onto \( X^+ \) as follows. Fix the point \( P_0 = (m, 0, 0, 0) \in X^+ \). The isotropy subgroup of \( P_0 \) is \( SU(2) \) so a diffeomorphism

\[
\beta_{P_0} : SL(2, \mathbb{C}) / SU(2) \rightarrow X^+ 
\]

is defined by

\[
\beta_{P_0}([A]) = A \cdot P_0
\]

for every \([A] \in SL(2, \mathbb{C}) / SU(2)\). Notice that this is independent of the representative \( A \) chosen for \([A]\) because \( SU(2) \) fixes \( P_0 \). The inverse diffeomorphism carries \( p \in X^+ \) to the unique \([A] \in SL(2, \mathbb{C}) / SU(2)\) with \( A \cdot P_0 = p \) for each representative \( A \) of \([A]\).

\[
(\beta_{P_0})^{-1}(p) = (\beta_{P_0})^{-1}(A \cdot P_0) = [A] \quad (2.60)
\]
We would like to make a smooth selection $\omega(p)$ of a representative of this $[A]$ for $p \in X_+^m$. Since the smooth principal SU(2)-bundle $SL(2, \mathbb{C}) \to SL(2, \mathbb{C})/SU(2) \cong \mathbb{R}^3$ is trivial it has smooth global sections. Choose such a section

$$u : SL(2, \mathbb{C})/SU(2) \to SL(2, \mathbb{C})$$

and define

$$\omega : X_+^m \to SL(2, \mathbb{C})$$

by

$$\omega(p) = (u \circ \beta_p^{-1})(p). \quad (2.61)$$

Then $\omega(p)$ is a smooth function from $X_+^m$ to $SL(2, \mathbb{C})$ satisfying

$$\omega(p) \cdot P_0 = p \quad (2.62)$$

for every $p \in X_+^m$.

**Remark 2.6.3.** The map $\omega : X_+^m \to SL(2, \mathbb{C})$ depends, of course, on the choice of the section $u : SL(2, \mathbb{C})/SU(2) \to SL(2, \mathbb{C})$. Somewhat later we will want to make a choice for $u$ that produces an $\omega$ that is equivariant with respect to certain SU(2)-actions on $X_+^m$ and $SL(2, \mathbb{C})$. We define actions of SU(2) on $SL(2, \mathbb{C})$, $SL(2, \mathbb{C})/SU(2)$, and $X_+^m$ as follows. Let $U \in SU(2)$ and define, for $A \in SL(2, \mathbb{C})$, $[A] \in SL(2, \mathbb{C})/SU(2)$, and $p = \beta_P([A]) \in X_+^m$,

$$U \cdot A = UAU^{-1}$$

$$U \cdot [A] = [UA]$$

$$U \cdot p = U \cdot \beta_P([A]) = \beta_P(U \cdot [A]) = \beta_P([UA]).$$

**Exercise 2.6.4.** Show that, with respect to these SU(2)-actions, $\omega$ is equivariant

$$\omega(U \cdot p) = U \cdot \omega(p)$$

if and only if $u$ is equivariant

$$u(U \cdot [A]) = U \cdot u([A]).$$

Thus, we need only construct a section $u : SL(2, \mathbb{C})/SU(2) \to SL(2, \mathbb{C})$ satisfying

$$u([UA]) = U \cdot u([A]) U^{-1}$$

for any $U \in SU(2)$. For this we use the fact that any $A \in SL(2, \mathbb{C})$ has a unique polar decomposition.
2.6 Momentum Space

\[ A = PV, \]

where \( P \) is positive definite Hermitian and \( V \) is in SU(2). Notice that \( \det P = 1 \) so \( P \) is in SL(2, \( \mathbb{C} \)) and, moreover,

\[ [A] = A SU(2) = (PV) SU(2) = PSU(2) = [P]. \]

Now define \( u : SL(2, \mathbb{C})/SU(2) \to SL(2, \mathbb{C}) \) by

\[ u([A]) = u([P]) = P. \]

Next we observe that

\[ [UA] = (UA) SU(2) = (UPV) SU(2) = (UP) SU(2) = (UPU^{-1}) SU(2) = [UPU^{-1}] \]

Consequently, \( u([UA]) = u([UPU^{-1}]) = P' \), where \( P'V' \) is the polar decomposition of \( UPU^{-1} \). But since \( P \) is positive definite Hermitian it can be written in the form \( P = e^B \), where \( B \) is Hermitian, and therefore

\[ UPU^{-1} = Ue^BU^{-1} = e^{UBU^{-1}} \]

which is also positive definite Hermitian. By uniqueness of the polar decomposition of \( UPU^{-1} \), \( P' = UPU^{-1} \) and \( V' = id_{2\times 2} \) so

\[ u([UA]) = UPU^{-1} = U u([A]) U^{-1} \]

as required. With this \( u, \omega \) is equivariant with respect to the SU(2)-actions on \( X_m^+ \) and SL(2, \( \mathbb{C} \)).

2.6.2 Invariant Measures

Since \( \mathcal{L}_+^+ \) acts transitively on \( X_m^+ \) one can ask if there is a Borel measure on \( X_m^+ \) that is invariant under this action in the same way that the Lebesgue measure on \( \mathbb{R}^n \) is invariant under the action of the translation group. Indeed, there is such a measure and, up to a positive multiple, it is unique. The uniqueness modulo positive constants is proved in the Appendix to Section IX.8 of [RS2]. We will show that one such measure \( \mu_m \) is defined as follows. For any Borel set \( B \subseteq X_m^+ \), \( \pi_+(B) \) is a Borel subset of \( \mathbb{R}^3 \), where

\[ \pi_+ : X_m^+ \to \mathbb{R}^3 \]

\[ \pi_+(p_0, p_1, p_2, p_3) = (p_1, p_2, p_3) \]

is the projection. Now define
\[ \mu_m(B) = \int_{\pi_+(B)} \frac{d^3p}{2\omega_p}, \]

where

\[ \omega_p = \sqrt{m^2 + \|p\|^2} = \sqrt{m^2 + p_1^2 + p_2^2 + p_3^2} \]

and \( d^3p = dp_1 dp_2 dp_3 \) denotes integration with respect to Lebesgue measure on \( \mathbb{R}^3 \). Notice that, if \( p = \pi_+(p) \), then \( \omega_p = p_0 \).

Remark 2.6.4. This definition is not as mysterious as it might seem, being just a special case of a general construction of surface measures induced on smooth hypersurfaces in \( \mathbb{R}^n \) by the Lebesgue measure on \( \mathbb{R}^n \) (see pages 78-79, Section IX.9, of [RS2] and also Section 1.2, Chapter III, of [GS]).

We will show that \( \mu_m \) is invariant under the action of \( \mathcal{L}^+_+ \) on \( X^+_m \) by simply applying the Change of Variables Formula which we record here.

**Theorem 2.6.1.** (Change of Variables Formula) Let \( U \) and \( V \) be open subsets of \( \mathbb{R}^n \) and \( g \) a \( C^1 \)-diffeomorphism of \( U \) onto \( V \) with Jacobian determinant \( \det(g') \). Suppose \( B \subseteq U \) is a Borel set and \( f : V \to \mathbb{R} \) is Lebesgue integrable. Then \( g(B) \subseteq V \) is a Borel set. \( (f \circ g) | \det(g') \) is integrable on \( B \) and

\[ \int_{g(B)} f d^n x = \int_B (f \circ g) | \det(g') | d^n x. \]

Denote by \( \Lambda \) also the diffeomorphism of \( X^+_m \) onto itself corresponding to the action of some element \( \Lambda \) of \( \mathcal{L}^+_+ \). Also let \( s_+ : \mathbb{R}^3 \to X^+_m \) denote the map \( s_+(p) = s_+(p_1, p_2, p_3) = (\omega_p, p_1, p_2, p_3) \). Thus, \( s_+ : \mathbb{R}^3 \to X^+_m \) and \( \pi_+ : X^+_m \to \mathbb{R}^3 \) are inverse diffeomorphisms and \( g = \pi_+ \circ \Lambda \circ s_+ \) is a diffeomorphism of \( \mathbb{R}^3 \) onto \( \mathbb{R}^3 \) that carries \( \pi_+(B) \) onto \( \pi_+(\Lambda(B)) \). We need to show that \( \mu_m(\Lambda(B)) = \mu_m(B) \). For this we take \( f(p) = \frac{1}{\omega_p^2} \). Notice that

\[ (f \circ g)(p) = \frac{1}{2\omega_{\pi_+(\Lambda(p))}}. \]

The remainder of the proof is simplified if we recall (Theorem 2.3.5) that every \( \Lambda \in \mathcal{L}^+_+ \) can be written as the composition of two rotations and a boost so it will be enough to prove \( \mu_m(\Lambda(B)) = \mu_m(B) \) for boosts and rotations separately. Furthermore, all of the boosts are treated in exactly the same way so it will suffice to consider

\[ \Lambda = \begin{pmatrix} \cosh \theta & 0 & 0 & \sinh \theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh \theta & 0 & 0 & \cosh \theta \end{pmatrix}. \]
2.6 Momentum Space

From this we obtain

\[ \Lambda(\omega_p, \mathbf{p}) = \Lambda(\omega_p, p_1, p_2, p_3) \]
\[ = (\omega_p \cosh \theta + p_3 \sinh \theta, p_1, p_2, \omega_p \sinh \theta + p_3 \cosh \theta) \]

and therefore

\[ \omega_{\pi, (\Lambda(\omega_p, \mathbf{p}))} = \omega_p \cosh \theta + p_3 \sinh \theta. \]

Now, for the Jacobian \( g' \) we write the coordinate transformation of \( \mathbb{R}^3 \) determined by \( g \) as

\[ (p_1, p_2, p_3) \rightarrow (p_1, p_2, \sqrt{m^2 + p_1^2 + p_2^2 + p_3^2 \sinh \theta + p_3 \cosh \theta}) \]

and then just compute the derivatives. The result is

\[ g' = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
\frac{p_3 \sinh \theta + \omega_p \cosh \theta}{\omega_p} & \frac{p_3 \sinh \theta + \omega_p \cosh \theta}{\omega_p} & \frac{\omega_{\pi, (\Lambda(\omega_p, \mathbf{p}))}}{\omega_p}
\end{pmatrix}. \]

Consequently,

\[ \det(g') = \frac{p_3 \sinh \theta + \omega_p \cosh \theta}{\omega_p} = \frac{\omega_{\pi, (\Lambda(\omega_p, \mathbf{p}))}}{\omega_p} \]

which is positive because \( \Lambda \) is orthochronous. Substituting all of this into the Change of Variables Formula (Theorem 2.6.1) gives

\[ \mu_m(\Lambda(B)) = \int_{\pi, (\Lambda(B))} \frac{d^3 \mathbf{p}}{2 \omega_p} = \int_{\pi, (B)} \frac{1}{2 \omega_{\pi, (\Lambda(\omega_p, \mathbf{p}))}} \frac{\omega_{\pi, (\Lambda(\omega_p, \mathbf{p}))}}{\omega_p} d^3 \mathbf{p} \]
\[ = \int_{\pi, (B)} \frac{d^3 \mathbf{p}}{2 \omega_p} \]
\[ = \mu_m(B) \]

as required.

Exercise 2.6.5. Complete the proof by showing that \( \mu_m(\Lambda(B)) = \mu_m(B) \) when \( \Lambda \) is in the rotation subgroup of \( \mathcal{L}_+^1 \).

Remark 2.6.5. The Hilbert space \( L^2(X_m^+, \mu_m) \) plays a particularly prominent role in the construction of a rigorous model of the so-called Wightman axioms for scalar quantum field theory. We will return to this somewhat later (also see Section X.7 of [RS2]).

Since the action of \( \text{SL}(2, \mathbb{C}) \) on \( X_m^+ \) is defined in terms of the \( \mathcal{L}_+^1 \)-action via the double covering \( \kappa : \text{SL}(2, \mathbb{C}) \rightarrow \mathcal{L}_+^1 \), the measure \( \mu_m \) is also invariant under the
Recalling the $X_m^\ast_\mu$ is diffeomorphic to the homogeneous manifold $\text{SL}(2, \mathbb{C})/\text{SU}(2)$ which admits an $\text{SL}(2, \mathbb{C})$-action defined by $B \cdot [A] = [BA]$ for all $B \in \text{SL}(2, \mathbb{C})$ and all $[A] \in \text{SL}(2, \mathbb{C})/\text{SU}(2)$, we would like to move $\mu_m$ to an $\text{SL}(2, \mathbb{C})$-invariant measure on $\text{SL}(2, \mathbb{C})/\text{SU}(2)$.

**Remark 2.6.6.** Recall that, if $(X, \mathcal{A}, m)$ is a measure space, $(Y, \mathcal{B})$ is a measurable space and $g : (X, \mathcal{A}) \to (Y, \mathcal{B})$ is a measurable function, then the **pushforward measure** $g_\ast(m)$ on $(Y, \mathcal{B})$ is defined by

$$
(g_\ast(m))(B) = m(g^{-1}(B)) \quad \forall B \in \mathcal{B}.
$$

There is an abstract change of variables formula (Theorem C, Section 39, of [Hal1]) that asserts the following. If $f$ is any extended real-valued measurable function on $(Y, \mathcal{B})$, then $f$ is integrable with respect to $g_\ast(m)$ if and only if $f \circ g$ is $m$-integrable and, in this case,

$$
\int_Y f(y) d(g_\ast(m))(y) = \int_X (f \circ g)(x) dm(x).
$$

Furthermore, (2.63) holds in the stronger sense that, if either side is defined (even if it is not finite), then the other side is defined and they agree.

**Exercise 2.6.6.** Let $\beta_{p_0}^{-1}$ be the diffeomorphism defined by (2.60). Show that

$$
\mu = (\beta_{p_0}^{-1})_\ast(\mu_m)
$$

is an $\text{SL}(2, \mathbb{C})$-invariant measure on $\text{SL}(2, \mathbb{C})/\text{SU}(2)$.

### 2.6.3 Momentum Space and the Character Group

The next item we require is the identification of $P^{1,3}_1$ with the character group $\hat{R}^{1,3}_1$ of the additive (translation) group of $R^{1,3}$ (character groups for Abelian Lie groups are reviewed in Remark 1.5.2). This will simplify the application of the Mackey machine in Section 2.8.

We begin by reviewing the relevant notation. Choose a fixed, but arbitrary admissible basis $\{e_0, e_1, e_2, e_3\}$ for Minkowski spacetime $M$ and thereby identify $M$ with $\mathbb{R}^{1,3}$, the elements of which will be denoted $x = (x^\mu) = (x^0, x^1, x^2, x^3) = (x, y) = (y^\mu) = (y^0, y^1, y^2, y^3) = (y, y)$. We will use the same symbol $\mathbb{R}^{1,3}$ for its additive group of translations. The dual basis for $P^{1,3}_1 = (\mathbb{R}^{1,3})^\ast$ is denoted $\{e^0, e^1, e^2, e^3\}$ and the elements of $P^{1,3}_1$ will generally be written in the dual basis as $p = (p_0, p_1, p_2, p_3) = (p_0, p)$. The double covering of the proper, orthochronous Lorentz group $\mathcal{L}_+^\ast$ is written...
2.6 Momentum Space

\[ \kappa : \text{SL}(2, \mathbb{C}) \rightarrow \mathcal{L}_+^1 \]

and \( \text{SL}(2, \mathbb{C}) \) acts on \( \mathbb{R}^{1,3} \) by defining

\[ \sigma : \text{SL}(2, \mathbb{C}) \rightarrow \text{GL}(\mathbb{R}^{1,3}) \]

by

\[ \sigma(A)(x) = A \cdot x = \kappa(A)x, \]

where \( A \in \text{SL}(2, \mathbb{C}) \) and \( \kappa(A)x \) is the matrix product of \( \kappa(A) \in \mathcal{L}_+^1 \) with the (column) vector \( x \in \mathbb{R}^{1,3} \). With this action we define inhomogeneous \( \text{SL}(2, \mathbb{C}) \), denoted ISL(2, \mathbb{C}), to be the semi-direct product of \( \mathbb{R}^{1,3} \) (thought of as the translation group) and \( \text{SL}(2, \mathbb{C}) \), that is,

\[ \text{ISL}(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes \text{SL}(2, \mathbb{C}). \]

As for \( P^1_+ \) one generally omits the explicit reference to \( \sigma \) in the notation and writes

\[ \text{ISL}(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes \text{SL}(2, \mathbb{C}). \]

The action of \( \text{SL}(2, \mathbb{C}) \) on \( \mathbb{R}^{1,3} \) induces a contragredient action

\[ \sigma^* : \text{SL}(2, \mathbb{C}) \rightarrow \text{GL}(\mathbb{R}^{1,3})^* \]

of \( \text{SL}(2, \mathbb{C}) \) on the vector space dual \( \mathbb{P}^{1,3} = (\mathbb{R}^{1,3})^* \) defined by

\[ [\sigma^*(A)(\gamma)](x) = [A \cdot \gamma](x) = \gamma(\sigma(A^{-1})(x)), \]

where \( A \in \text{SL}(2, \mathbb{C}) \), \( \gamma : \mathbb{R}^{1,3} \rightarrow \mathbb{R} \) is in \( \mathbb{P}^{1,3} = (\mathbb{R}^{1,3})^* \), and \( x \in \mathbb{R}^{1,3} \). In terms of components relative to \( \{e^0, e^1, e^2, e^3\} \)

\[ \sigma^*(A)(p) = A \cdot p = \kappa(A^{-1})^T p, \]

where \( T \) means transpose and \( \kappa(A^{-1})^T p \) is the matrix product of \( \kappa(A^{-1})^T \) and the (column) vector \( p \).

Next we consider the character group \( \hat{\mathbb{R}}^{1,3} \) of the translation group \( \mathbb{R}^{1,3} \) (see Remark 1.5.2). This has nothing to do with the signature of the inner product so \( \hat{\mathbb{R}}^{1,3} \) is the same as \( \hat{\mathbb{R}}^4 \). The action \( \sigma \) of \( \text{SL}(2, \mathbb{C}) \) on \( \mathbb{R}^{1,3} \) also induces an action \( \hat{\sigma} \) of \( \text{SL}(2, \mathbb{C}) \) on \( \hat{\mathbb{R}}^{1,3} \) as follows. Any \( \xi \in \hat{\mathbb{R}}^{1,3} \) is a continuous homomorphism from the additive Abelian group \( \mathbb{R}^{1,3} \) of translations to the multiplicative group \( S^1 \) of complex numbers of modulus one. For any \( A \in \text{SL}(2, \mathbb{C}) \) we define

\[ [\hat{\sigma}(A)(\xi)](a) = [A \cdot \xi](a) = \xi(\sigma(A^{-1})(a)) \]

for every \( a \in \mathbb{R}^{1,3} \). In Remark 1.5.2 it is shown that every \( \xi \) in \( \hat{\mathbb{R}}^{1,3} \) can be written uniquely in the form
\[ e^{i(p_0 x^0 + p_1 x^1 + p_2 x^2 + p_3 x^3)} \]

for some \((p_0, p_1, p_2, p_3) \in \mathbb{R}^4\) so that

\[(p_0, p_1, p_2, p_3) \in \mathbb{R}^4 \mapsto e^{i(p_0 x^0 + p_1 x^1 + p_2 x^2 + p_3 x^3)} \in \hat{\mathbb{R}}^{1,3}\]

is a group isomorphism of the additive group \(\mathbb{R}^4\) onto the character group \(\hat{\mathbb{R}}^{1,3}\).

**Remark 2.6.7.** It will be convenient later on to note that the characters in \(\hat{\mathbb{R}}^{1,3}\) can equally well be written uniquely as

\[ e^{i(p_0 x^0 + p_1 x^1 + p_2 x^2 + p_3 x^3)} \]

simply by changing the signs of \(p_1, p_2\) and \(p_3\).

We would like to show that \(\hat{\mathbb{R}}^{1,3}\) admits the structure of a real vector space and that with this structure there is a linear isomorphism of \((\hat{\mathbb{R}}^{1,3})^*\) onto \(\hat{\mathbb{R}}^{1,3}\) that intertwines the actions of \(\text{SL}(2, \mathbb{C})\) on \(\hat{\mathbb{R}}^{1,3}\) and \((\mathbb{R}^{1,3})^*\).

**Exercise 2.6.7.** Define addition \(\boxplus\) and real scalar multiplication on \(\hat{\mathbb{R}}^{1,3}\) by

\[ (\xi_1 \boxplus \xi_2)(a) = \xi_1(a)\xi_2(a) \]
\[ (c\xi)(a) = \xi(ca) \quad \forall c \in \mathbb{R} \]

and show that, with these operations, \(\hat{\mathbb{R}}^{1,3}\) becomes a real vector space.

**Exercise 2.6.8.** Let \(\phi : \mathbb{R} \to S^1\) be any nontrivial character of the additive group \(\mathbb{R}\), say, \(\phi(x) = e^{ipx}\), where \(p\) is a nonzero real number. Now define \(\varphi : (\mathbb{R}^{1,3})^* \to \hat{\mathbb{R}}^{1,3}\) by

\[ \varphi(\gamma) = \phi \circ \gamma \]

for all \(\gamma\) in \((\mathbb{R}^{1,3})^*\). Show that \(\varphi\) is a vector space isomorphism. **Hint:** For surjectivity note that \(\phi : \mathbb{R} \to S^1\) is a covering space with \(\phi(0) = 1\) and that \(\mathbb{R}^{1,3}\) is simply connected. It follows that any element of \(\hat{\mathbb{R}}^{1,3}\) has a unique lift \(\gamma : \mathbb{R}^{1,3} \to \mathbb{R}\) with \(\gamma(0) = 0\) (see, for example, Theorem 5.1 and Theorem 6.1 of [Gre]). Now show that \(\gamma\) is linear.

**Exercise 2.6.9.** Show that, for every \(A\) in \(\text{SL}(2, \mathbb{C})\),

\[ \hat{\sigma}(A) = \varphi \circ \sigma^*(A) \circ \varphi^{-1} \]

With this we conclude that \(\hat{\mathbb{R}}^{1,3}\) can be identified with the vector space dual \((\mathbb{R}^{1,3})^*\) of \(\mathbb{R}^{1,3}\), that is, with momentum space \(\mathbb{P}^{1,3}\), in such a way that the induced
2.7 Spacetime Symmetries and Projective Representations

Symmetries in classical mechanics were discussed in Section A.2 (Lagrangian picture) and Section A.3 (Hamiltonian picture). The corresponding notion in quantum mechanics was introduced in Section A.4 (see page 154). We recall that a symmetry of a quantum system with Hilbert space $\mathcal{H}$ is a bijection of the state space $\mathcal{P}(\mathcal{H})$ onto itself that preserves transition probabilities. By Wigner’s Theorem (Theorem 1.3.1 and Theorem A.4.2), each of these arises from an operator on $\mathcal{H}$ that is either unitary or anti-unitary. Our objective in this section is to take the first step toward “relativistic quantum mechanics” by defining what it means for a quantum system to be “relativistically invariant”. To put the matter simply we will take this to mean that the Poincaré group $\mathcal{P}^+$ “acts by symmetries” on the state space $\mathcal{P}(\mathcal{H})$ of the quantum system. But to say that $\mathcal{P}^+$ acts by symmetries on $\mathcal{P}(\mathcal{H})$ means simply that there is a projective representation of $\mathcal{P}^+$ on $\mathcal{H}$, that is, a continuous group homomorphism $\varphi : \mathcal{P}^+ \to \text{Aut}(\mathcal{P}(\mathcal{H}))$ of $\mathcal{P}^+$ into the automorphism group of $\mathcal{P}(\mathcal{H})$ (again, see Section 1.3). Different “types” of relativistic invariance correspond to different projective representations just as different types of rotational invariance (scalar, vector, tensor) correspond to different representations of $\text{SO}(3)$ in classical mechanics. It would be nice then to find all of the (irreducible) projective representations of $\mathcal{P}^+$ on $\mathcal{H}$. In this section we will simply sketch how this can be reduced to the well-studied problem of finding the irreducible unitary representations of the double cover $\text{ISL}(2, \mathbb{C})$ of $\mathcal{P}^+$ on $\mathcal{H}$.

The principal tool we require is a highly nontrivial result of Bargmann which asserts that every projective representation of $\text{ISL}(2, \mathbb{C})$ on $\mathcal{H}$ lifts to a unitary representation of $\text{ISL}(2, \mathbb{C})$ on $\mathcal{H}$. Such liftings of projective representations do not exist in general and Bargmann’s Theorem depends crucially on the topology of $\text{ISL}(2, \mathbb{C})$ (see Section 1.3). For the proof of the following result we refer to Theorem 14.3 of [VDB].

Theorem 2.7.1. (Bargmann’s Theorem) Let $\mathcal{H}$ be a complex, separable Hilbert space. Every projective representation $\varphi : \text{ISL}(2, \mathbb{C}) \to \text{Aut}(\mathcal{P}(\mathcal{H}))$ of $\text{ISL}(2, \mathbb{C})$ on $\mathcal{H}$ lifts to a unique unitary representation $\tilde{\varphi} : \text{ISL}(2, \mathbb{C}) \to \mathcal{U}(\mathcal{H})$ of $\text{ISL}(2, \mathbb{C})$ on $\mathcal{H}$.

\[
\begin{array}{c}
\mathcal{U}(\mathcal{H}) \\
\uparrow \\
\text{ISL}(2, \mathbb{C}) \to \text{Aut}(\mathcal{P}(\mathcal{H}))
\end{array}
\]
Here the downward vertical arrow is the restriction to \( \mathcal{H} \) of the quotient map
\[ Q : \mathcal{H} = \mathcal{H} \cup \mathcal{H}^+ \to Aut(\mathcal{P}(\mathcal{H})) \] described in Proposition 1.3.2 so
\[ \rho = Q \circ \tilde{\rho}. \]

Moreover, \( \rho \) is irreducible if and only if \( \tilde{\rho} \) is irreducible.

The following is Corollary 14.4 of [VDB], but the proof is simple and a nice application of Schur’s Lemma. Since we will need to see the specific construction employed in the proof we will record it here as well.

**Theorem 2.7.2.** Let \( \mathcal{H} \) be a complex, separable Hilbert space. Every irreducible, unitary representation \( \tilde{\rho} : ISL(2, \mathbb{C}) \to Aut(\mathcal{H}) \) of ISL(2, \( \mathbb{C} \)) on \( \mathcal{H} \) naturally induces an irreducible, projective representation \( \tilde{\rho} : \mathcal{P}^+_1 \to Aut(\mathcal{P}(\mathcal{H})) \) of the Poincaré group \( \mathcal{P}^+_1 \) on \( \mathcal{H} \).

**Proof.** We have already seen that \( \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \) is the universal cover of \( \mathbb{R}^{1,3} \rtimes \mathcal{L}^+_1 \), that the kernel of the covering map \( \kappa' = id_{\mathbb{R}^{1,3}} \times \kappa : \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \to \mathbb{R}^{1,3} \rtimes \mathcal{L}^+_1 \) is \( \mathbb{Z}_2 = \{ 0, \pm \text{id}_{\mathbb{R}^{1,3}} \} \) and that this kernel is precisely the center of \( \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \). Now let \( \tilde{\rho} : \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \to Aut(\mathcal{H}) \) be an irreducible, unitary representation of \( \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \) on a complex, separable Hilbert space \( \mathcal{H} \). It follows from Schur’s Lemma 1.2.1 that \( \tilde{\rho} \) restricted to \( Kernel(\kappa') \) acts by scalars of modulus one. It is shown in Section 1.3 that \( \tilde{\rho} \) induces an irreducible, projective representation \( \rho : \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \to Aut(\mathcal{P}(\mathcal{H})) \) defined by
\[ \rho(x,A)(|\psi\rangle) = [\tilde{\rho}(x,A)|\psi\rangle]. \]

Since \( \tilde{\rho} \) acts by scalars of modulus one on \( Kernel(\kappa') \), the projective representation \( \rho \) is trivial on \( Kernel(\kappa') \). Consequently, factoring by \( Kernel(\kappa') \) gives an irreducible, projective representation \( \rho \) of the quotient, that is, of the Poincaré group \( \mathcal{P}^+_1 = ISL(2, \mathbb{C})/Kernel(\kappa') \). \( \square \)

The objects of interest in relativistic quantum mechanics are the irreducible, projective representations of the Poincaré group \( \mathcal{P}^+_1 = \mathbb{R}^{1,3} \rtimes \mathcal{L}^+_1 \) (see Remark A.4.2). We will now show that the previous two theorems reduce the problem of finding these to that of enumerating the irreducible, unitary representations of \( ISL(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \). Theorem 2.7.2 assures us that any irreducible, unitary representation of \( ISL(2, \mathbb{C}) \) induces an irreducible, projective representation of \( \mathcal{P}^+_1 \), so we need only show that every irreducible, projective representation \( \rho \) of \( \mathcal{P}^+_1 \) gives rise to an irreducible, unitary representation of \( ISL(2, \mathbb{C}) \) that induces \( \rho \) in the sense of Theorem 2.7.2. Suppose then that \( \rho : \mathbb{R}^{1,3} \rtimes \mathcal{L}^+_1 \to Aut(\mathcal{P}(\mathcal{H})) \) is an irreducible, projective representation of \( \mathcal{P}^+_1 \). Composition with the double cover homomorphism \( \mathbb{R}^{1,3} \rtimes SL(2, \mathbb{C}) \to \mathbb{R}^{1,3} \rtimes \mathcal{L}^+_1 \) gives an irreducible, projective representation \( \tilde{\rho} \) of \( ISL(2, \mathbb{C}) \). Bargmann’s Theorem 2.7.1 then implies that \( \rho \) has a unique lift to an irreducible, unitary representation \( \tilde{\rho} \) of \( ISL(2, \mathbb{C}) \).
Exercise 2.7.1. Show that \( \hat{\rho} \) induces \( \rho \) in the sense of Theorem 2.7.2. Hint: Follow the construction in the proof of Theorem 2.7.2.

Our problem then is to enumerate the irreducible, unitary representations of ISL(2, \( \mathbb{C} \)) and for this we can apply the “Mackey machine” described in Section 1.5. In the next section we will describe that part of enumeration that is of interest to us here (the so-called positive energy representations with mass \( m > 0 \)).

2.8 Positive Energy Representations of \( \mathcal{D}_+^1 \) with \( m > 0 \)

The Mackey machine is described in Section 1.5 and we will simply follow the instructions enumerated on page 41 when \( G = \text{ISL}(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes \text{SL}(2, \mathbb{C}) \). The first of these requires that we select an orbit of the SL(2, \( \mathbb{C} \))-action on the character group \( \mathbb{R}^{1,3} \) and a point \( P_0 \) in it. In Section 2.3 we identified \( \mathbb{R}^{1,3} \) with the vector space dual \( \mathbb{P}^{1,3} = (\mathbb{R}^{1,3})^* \) of \( \mathbb{R}^{1,3} \) and the SL(2, \( \mathbb{C} \))-action on \( \mathbb{R}^{1,3} \) with the contragredient action of SL(2, \( \mathbb{C} \)) on \( \mathbb{P}^{1,3} \), that is,

\[
A \cdot p = \kappa(A^{-1})^T p,
\]

where \( \kappa : \text{SL}(2, \mathbb{C}) \to \mathcal{L}_+^1 \) is the covering map (Section 2.4). It will therefore suffice to find the orbits of this action on \( \mathbb{P}^{1,3} \). Since \( \kappa(A^{-1})^T \) is in \( \mathcal{L}_+^1 \), these are the same as the \( \mathcal{L}_+^1 \)-orbits in momentum space and these were described in Section 2.4. Notice that these are all closed subsets of \( \mathbb{P}^{1,3} \) so ISL(2, \( \mathbb{C} \)) is a regular semi-direct product. For our purposes we will be interested only in the mass hyperboloids \( X_m^+, m > 0 \).

Remark 2.8.1. We have chosen to restrict our attention to \( X_m^+ \) for reasons that are in the physics, not the mathematics (see Remark 2.6.1). We will eventually describe the physical interpretation of the results we derive here for \( X_m^+ \), but for the remaining orbits there are subtleties such as negative energy states that we prefer not to become entangled in at the moment. For those who would like to see the full story we refer to [Simms] and [Vara].

We have shown in Section 2.4 that, for any point \( p \in X_m^+ \), the isotropy subgroup of \( p \) for the \( \mathcal{L}_+^1 \)-action on \( X_m^+ \) is isomorphic to the rotation subgroup \( R \cong \text{SO}(3) \) of \( \mathcal{L}_+^1 \). Consequently, the isotropy subgroup for the SL(2, \( \mathbb{C} \))-action is the pre-image under \( \kappa \) of \( R \) and this is isomorphic to SU(2) (Section 2.4).

Remark 2.8.2. Recall that every point \( p \in X_m^+ \) gives rise to a character \( \xi_p \in \mathbb{R}^{1,3} \) defined by

\[
p = (p_0, p_1, p_2, p_3) \in X_m^+ \mapsto \xi_p(x) = e^{i(p_0x^0 - p_1x^1 - p_2x^2 - p_3x^3)}.
\]
For the remainder of the discussion we will fix the “base point” \( P_0 = (m, 0, 0, 0) \in X^*_m \) and the corresponding character

\[
\xi_0 = e^{imx_0} \tag{2.64}
\]

in \( \mathbb{R}^{1,3} \) so that by SU(2) we mean the subgroup of SL(2, C) that leaves \( \xi_0 \) fixed in \( \mathbb{R}^{1,3} \) and leaves \( P_0 \) fixed in \( X^+_m \).

We have also shown in Section 2.4 that \( X^+_m \) admits a measure \( \mu_m \) that is invariant under the \( \mathcal{L}_t \)-action and therefore also under the SL(2, C)-action. Recall that it is defined in the following way. For any Borel set \( B \subseteq X^+_m \), the projection \( \pi_t(B) \) is a Borel subset of \( \mathbb{R}^3 \) and we define

\[
\mu_m(B) = \int_{\pi_t(B)} \frac{d^3p}{2\omega_p} = \int_{\pi_t(B)} \frac{d^3p}{2\sqrt{m^2 + ||p||^2}},
\]

where \( d^3p = dp_1 dp_2 dp_3 \) denotes integration with respect to Lebesgue measure on \( \mathbb{R}^3 \). The pushforward measure

\[
\mu = (\xi_0^{-1}).(\mu_m)
\]

is then an SL(2, C)-invariant measure on SL(2, C)/SU(2) (see (2.60) and Remark 2.6.6). This fulfills the second requirement of the Mackey machine (see page 41).

To set the Mackey machine in motion for ISL(2, C) we must now select a strongly continuous, irreducible, unitary representation of the isotropy group SU(2). In Example 1.2.1 we saw that these are precisely the spinor representations

\[
\mathcal{D}(j/2) : SU(2) \to \mathcal{U}(\mathcal{C}(\mathcal{D}(j/2))),
\]

where \( j \geq 0 \) is an integer and \( \mathcal{C}(\mathcal{D}(j/2)) \) is the \((j + 1)\)-dimensional Hilbert space of carriers of \( \mathcal{D}(j/2) \), that is, the space of \( 2^j \)-tuples \( \xi^{A_1 A_2 \cdots A_j} A_1, A_2, \ldots, A_j = 1, 2 \), of complex numbers that are symmetric under permutations of \( A_1 A_2 \cdots A_j \). We now fix one of these representations \( \mathcal{D}(j/2) \) for some \( j \geq 0 \).

Next we are to consider the principal SU(2)-bundle

\[
\pi : SL(2, \mathbb{C}) \to SL(2, \mathbb{C})/SU(2) \cong X^*_m \cong \mathbb{R}^3,
\]

where the right action of SU(2) on SL(2, C) is right multiplication. This bundle is trivial because any principal bundle over any \( \mathbb{R}^n \) is trivial and it follows that the vector bundle

\[
\pi_{\mathcal{D}(j/2)} : SL(2, \mathbb{C}) \times_{\mathcal{D}(j/2)} \mathcal{C}(\mathcal{D}(j/2)) \to SL(2, \mathbb{C})/SU(2) \cong X^*_m \cong \mathbb{R}^3,
\]
associated to it by $\mathcal{D}^{(j/2)}$ is also trivial. We have seen that the Hilbert space $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ of sections of this vector bundle that are $L^2$ with respect to the invariant measure $\mu$ on $\text{SL}(2, \mathbb{C})/\text{SU}(2)$ can be identified with equivalence classes (modulo equality up to sets of measure zero) of functions $f : \text{SL}(2, \mathbb{C}) \to \mathbb{C}(\mathcal{D}^{(j/2)})$ that are measurable with respect to Haar measure on $\text{SL}(2, \mathbb{C})$ and satisfy each of the following.

1. $f(AU) = \mathcal{D}^{(j/2)}(U^{-1})(f(A))$ for all $U \in \text{SU}(2)$ and for almost all $A \in \text{SL}(2, \mathbb{C})$.

2. $\int_{\text{SL}(2, \mathbb{C})/\text{SU}(2)} \|f([A])\|^2 d\mu([A]) < \infty$, where $\|f([A])\| = \|f(A)\|_{\mathbb{C}(\mathcal{D}^{(j/2)})}$.

For the moment we will think of $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ as the space of such functions and then will switch back to sections when the need arises.

**Exercise 2.8.1.** Show that

$$\int_{\text{SL}(2, \mathbb{C})/\text{SU}(2)} \|f([A])\|^2 d\mu([A]) = \int_{X_+} \|f(\omega(p))\|^2_{\mathbb{C}(\mathcal{D}^{(j/2)})} d\mu_m(p).$$

**Hint:** Remark 2.6.6.

For the next step in the Mackey procedure we recall (Section 1.4) that the representation $\mathcal{D}^{(j/2)}$ of $\text{SU}(2)$ on $\mathbb{C}(\mathcal{D}^{(j/2)})$ induces a unitary representation of $\text{SL}(2, \mathbb{C})$ on $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ which sends any $A \in \text{SL}(2, \mathbb{C})$ to the operator on $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ which left translates an $f \in \mathcal{H}^{\mathcal{D}^{(j/2)}}$ by $A$, that is,

$$\text{Ind}^{\mathcal{H}}_{\text{SU}(2)}(\mathcal{D}^{(j/2)}) : \text{SL}(2, \mathbb{C}) \to \mathcal{H}^{\mathcal{D}^{(j/2)}}$$

is defined by

$$\left[ \text{Ind}^{\mathcal{H}}_{\text{SU}(2)}(\mathcal{D}^{(j/2)})(A) \right] f(A') = f(A^{-1}A'). \tag{2.65}$$

When the context makes the intention clear we will try to relieve some of the notational clutter by writing the action of $A$ in $\text{SL}(2, \mathbb{C})$ on $f$ corresponding to $\text{Ind}^{\mathcal{H}}_{\text{SU}(2)}(\mathcal{D}^{(j/2)})$ the way we write every other group action, that is, we will write (2.65) simply as

$$(A \cdot f)(A') = f(A^{-1}A'). \tag{2.66}$$

At this point we have induced a representation of $\text{SL}(2, \mathbb{C})$ on $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ and now we handle the first factor of $\text{ISL}(2, \mathbb{C}) = \mathbb{R}^{1,3} \rtimes \text{SL}(2, \mathbb{C})$. Then we will put them together to get a representation of $\text{ISL}(2, \mathbb{C})$ itself. For every $x$ in the additive group $\mathbb{R}^{1,3}$ we define a unitary multiplication operator $U(x)$ on $\mathcal{H}^{\mathcal{D}^{(j/2)}}$ by

$$[U(x)f](A') = [(A' \cdot \xi_0)(x)] f(A').$$
where $\xi_0$ is the character in the SL(2, C)-orbit of $\mathbb{R}^{1,3}$ corresponding to the base point $P_0 = (m, 0, 0, 0)$ in $X^*_m$ (see (2.64)). Note that as $A'$ varies over all of SL(2, C), $A' \cdot \xi_0$ varies over this entire orbit.

Finally, we put the two factors together and define, for every $(x, A) \in \text{ISL}(2, \mathbb{C})$ the unitary operator

$$U(x, A) = U(x) \circ \text{Ind}_{\text{SU}(2)}^{\text{SL}(2, \mathbb{C})}((\mathbb{D}^{j/2}))(A).$$

Thus,

$$[U(x, A)f](A') = [(A' \cdot \xi_0)(x)][A \cdot f](A') = [(A' \cdot \xi_0)(x)] f(A^{-1}A'). \quad (2.67)$$

Technically, this completes the application of the Mackey machine to ISL(2, C), but we will need to work a bit to put (2.67) into a usable form.

Notice that, since $A' \cdot \xi_0$ is a character, $(A' \cdot \xi_0)(x)$ is a complex number of modulus 1 for each $A' \in \text{SL}(2, \mathbb{C})$ and each $x \in \mathbb{R}^{1,3}$. We will now write out this phase factor explicitly. For this we recall that $(A' \cdot \xi_0)(x) = \xi_0((A')^{-1} \cdot x) = \xi_0((\kappa(A'))^{-1}x)$.

**Exercise 2.8.2.** Denote $\kappa(A')$ by $\Lambda \in \mathcal{L}^1$ and show that the $0^\text{th}$-component of $\kappa(A')^{-1}x$ is

$$\Lambda^0_0 x^0 - \Lambda^1_0 x^1 - \Lambda^2_0 x^2 - \Lambda^3_0 x^3.$$ 

Conclude that

$$(A' \cdot \xi_0)(x) = e^{i\kappa(A')(0^0x^0,A'_0x^1,A'_0x^2,A'_0x^3)}.$$ 

Now notice that, since $\Lambda = \kappa(A') \in \mathcal{L}^1$, $p = (p^0, p^1, p^2, p^3) = (mA^0_0, mA^1_0, mA^2_0, mA^3_0)$ satisfies $(p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 = m^2$ and so, if $p_a = \eta_{ab} p^b$, then $(p_0, p_1, p_2, p_3)$ is in $X^*_m$ for every $A' \in \text{SL}(2, \mathbb{C})$. Moreover, as $A'$ varies over all of SL(2, C), these points cover all of $X^*_m$. Consequently,

$$(A' \cdot \xi_0)(x) = e^{i(p^0 x^0 - p^1 x^1 - p^2 x^2 - p^3 x^3)} = e^{i(p \cdot x)} = e^{i(p \cdot x)},$$ 

where $p$ is just $m$ times the $0^\text{th}$-column of $\kappa(A')$. With this (2.67) becomes

$$[U(x, A)f](A') = e^{i(p \cdot x)} f(A^{-1}A'). \quad (2.68)$$

Next we will think a bit more about the second factor $f(A^{-1}A')$ and return to the view of $\mathcal{H}(\mathbb{D}^{j/2})$ as sections of a vector bundle. Recall that $f: \text{SL}(2, \mathbb{C}) \to \mathcal{C}(\mathbb{D}^{j/2})$ is a map from SL(2, C) to $\mathcal{C}(\mathbb{D}^{j/2})$ that is equivariant with respect to the SU(2)-actions, that is, satisfies

$$f(AU) = \mathbb{D}^{j/2}(U^{-1})(f(A)).$$
for all $U \in \text{SU}(2)$ and for almost all $A \in \text{SL}(2, \mathbb{C})$. As such, it determines a section
\[ s_f : \text{SL}(2, \mathbb{C})/\text{SU}(2) \to \text{SL}(2, \mathbb{C}) \times \mathcal{D}^{(j/2)} \mathbb{C} \]
of the vector bundle
\[ \pi_{\mathcal{D}^{(j/2)}} : \text{SL}(2, \mathbb{C}) \times \mathcal{D}^{(j/2)} \mathbb{C} \to \text{SL}(2, \mathbb{C})/\text{SU}(2) \]
given by
\[ s_f([A]) = [A', f(A')] \]
for all $[A'] \in \text{SL}(2, \mathbb{C})/\text{SU}(2)$ and any $A' \in [A']$ (see Remark 1.4.1). Moreover, any section $s$ of this vector bundle is $s_f$ for one and only one such equivariant map $f$.

Notice that both the base $\text{SL}(2, \mathbb{C})/\text{SU}(2)$ and the vector bundle space $\text{SL}(2, \mathbb{C}) \times \mathcal{D}^{(j/2)} \mathbb{C}$ admit continuous $\text{SL}(2, \mathbb{C})$-actions given by
\[ A \cdot [B] = [AB] \]
and
\[ A \cdot [B, v] = [AB, v]. \]
respectively, and that these commute with the projection $\pi_{\mathcal{D}^{(j/2)}}$. Now, we define an $\text{SL}(2, \mathbb{C})$-action on the sections in $\mathcal{E}_{\mathcal{D}^{(j/2)}}$ as follows. For any section $s \in \mathcal{E}_{\mathcal{D}^{(j/2)}}$ and any $A \in \text{SL}(2, \mathbb{C})$ we take $A \cdot s$ to be the section defined by
\[ (A \cdot s)([A']) = A \cdot (s(A^{-1} \cdot [A'])) = A \cdot s([A^{-1}A']) \]
for all $[A'] \in \text{SL}(2, \mathbb{C})/\text{SU}(2)$. The motivation for the definition is as follows. If $s_f$ is the section corresponding to $f$, then $A \cdot s$ is the section corresponding to $f \circ L_A$, where $L_A$ is the diffeomorphism of $\text{SL}(2, \mathbb{C})$ onto itself defined by $L_A(A') = A^{-1}A'$, that is,
\[ (A \cdot s_f)([A']) = s_{f \circ L_A}([A']) = [A', f(A^{-1}A')] \]
for all $A' \in \text{SL}(2, \mathbb{C})$ (compare (2.68)).

Since $\mu$ is an invariant measure on $\text{SL}(2, \mathbb{C})/\text{SU}(2)$ this action of $\text{SL}(2, \mathbb{C})$ on the sections in $\mathcal{E}_{\mathcal{D}^{(j/2)}}$ determines a unitary representation of $\text{SL}(2, \mathbb{C})$ on $\mathcal{E}_{\mathcal{D}^{(j/2)}}$ and one can show that this representation is strongly continuous. Note that we can write this in terms of the induced action of $\text{SL}(2, \mathbb{C})$ on the corresponding equivariant function $f$ as
\[ (A \cdot s_f)([A']) = [A', (A \cdot f)(A')]. \quad (2.69) \]
As an operator on sections the representation $U(x, A)$ therefore takes the form
where we again point out that $p = p(A')$ is $m$ times the $0^\text{th}$ column of $\kappa(A')$. This is looking more like the expression we want for $U(x,A)$, but there is still work to do.

The next thing to observe about the vector bundle $\pi_{2Dj/2} : \text{SL}(2,\mathbb{C}) \times_{2Dj/2} \mathcal{C}(\mathbb{D}^{j/2}) \to \text{SL}(2,\mathbb{C})/\text{SU}(2)$ is that, since $\text{SL}(2,\mathbb{C})/\text{SU}(2) \cong \mathbb{R}^3$, it is trivial so that any section $s$ of it can be identified with a function on $\text{SL}(2,\mathbb{C})/\text{SU}(2)$ whose value at any $[A]$ is a point in the fiber $\pi_{2Dj/2}^{-1}([A])$ above $[A]$. Furthermore, each of these fibers can be identified with the finite-dimensional Hilbert space $\mathcal{C}(\mathbb{D}^{j/2})$. Such an identification is obtained in the following way. The principal bundle $\pi : \text{SL}(2,\mathbb{C}) \to \text{SL}(2,\mathbb{C})/\text{SU}(2)$ is trivial for the same reason so we can select some global section $u : \text{SL}(2,\mathbb{C})/\text{SU}(2) \to \text{SL}(2,\mathbb{C})$. Then, for each $[A] \in \text{SL}(2,\mathbb{C})/\text{SU}(2)$, $u([A]) \in \text{SL}(2,\mathbb{C})$ so there is a unique $\psi([A]) \in \mathcal{C}(\mathbb{D}^{j/2})$ for which

$$s([A]) = [u([A]), \psi([A])].$$

Specifically, if $s = s_f$ corresponds to the equivariant map $f : \text{SL}(2,\mathbb{C}) \to \mathcal{C}(\mathbb{D}^{j/2})$, then $\psi = \psi_f = f \circ u$. Thus, given $u$, we can identify the section $s$ with the function $\psi : \text{SL}(2,\mathbb{C})/\text{SU}(2) \to \mathcal{C}(\mathbb{D}^{j/2})$. However, this identification of $s$ with a function from $\text{SL}(2,\mathbb{C})/\text{SU}(2)$ to $\mathcal{C}(\mathbb{D}^{j/2})$ is not unique since it depends on the choice of the section $u$. Indeed, if $\hat{u} : \text{SL}(2,\mathbb{C})/\text{SU}(2) \to \text{SL}(2,\mathbb{C})$ is another global section, then for each $[A] \in \text{SL}(2,\mathbb{C})/\text{SU}(2)$ we can write $\hat{u}([A]) = u([A])U([A])$ for some $U([A]) \in \text{SU}(2)$ and then, by definition of the equivalence relation defining the points of $\text{SL}(2,\mathbb{C}) \times_{2Dj/2} \mathcal{C}(\mathbb{D}^{j/2})$, $[\hat{u}([A]), U([A]) \cdot \psi([A])] = [u([A]), \psi([A])]$ so

$$s([A]) = [\hat{u}([A]), U([A]) \cdot \psi([A])].$$

In other words, $\psi$ is determined only up to the action of $\text{SU}(2)$ given by the representation $\mathbb{D}^{j/2}$.

Next we will need a few computational formulas. First we note that, when $s = s_f$, $f$ can be recovered from $\psi$. To see this we write

$$f(A') = f(u([A'])) u([A'])^{-1} A'$$

$$= \mathbb{D}^{j/2}(u([A']))^{-1} u([A']) f(u([A']))$$

so

$$f(A') = \mathbb{D}^{j/2}(u([A'])^{-1} u([A'])) \psi([A']).$$

\textit{Remark 2.8.3.} Notice that $u([A'])^{-1} A'$ is in $\text{SU}(2)$ because its projection into $\text{SL}(2,\mathbb{C})/\text{SU}(2)$ is the identity.
From this we obtain
\[
f(A^{-1}A') = D^{(j/2)}((A^{-1}A')^{-1}u([A^{-1}A']))\psi([A^{-1}A'])
\]
\[
= D^{(j/2)}(A')^{-1}Au(A^{-1} \cdot [A'])\psi(A^{-1} \cdot [A'])
\]
\[
= D^{(j/2)}((A')^{-1}u([A']))(A^{-1})^{-1}Au(A^{-1} \cdot [A'])\psi(A^{-1} \cdot [A'])
\]
\[
= D^{(j/2)}(A')^{-1}u([A'])D^{(j/2)}(u([A']))^{-1}Au(A^{-1} \cdot [A'])\psi(A^{-1} \cdot [A'])
\]
(2.72)

We have not yet defined an action of SL(2, C) on \(\psi\), but (2.71) suggests that we define \(A \cdot \psi\) in such a way that
\[
(A \cdot f)(A') = D^{(j/2)}((A')^{-1}u([A']))(A \cdot \psi)([A'])
\]
and (2.72) would then give
\[
(A \cdot \psi)([A']) = D^{(j/2)}((A')^{-1}u(A^{-1} \cdot [A']))\psi(A^{-1} \cdot [A']).
\]
(2.73)

Thus,
\[
f(A^{-1}A') = D^{(j/2)}((A')^{-1}u([A']))((A \cdot \psi)([A']))
\]
\[
= ((A')^{-1}u([A'])) \cdot ((A \cdot \psi)([A'])).
\]

We claim that, with the SL(2, C)-action (2.73),
\[
(A \cdot s)([A']) = [u([A']), (A \cdot \psi)([A'])]
\]
so that \(A \cdot \psi\) is the function associated with the section \(A \cdot s\). To see this we write \(s\) uniquely as \(s = s_f\) and compute
\[
(A \cdot s_f)([A']) = [A', f(A^{-1}A')]
\]
\[
= [A', ((A')^{-1}u([A'])) \cdot ((A \cdot \psi)([A']))]
\]
\[
= [A' \cdot ((A')^{-1}u([A']))(A \cdot \psi)([A'])]
\]
\[
= [u([A']), (A \cdot \psi)([A'])].
\]

As an operator on \(\psi\) the representation \(U(x, A)\) therefore takes the form
\[
[U(x, A)\psi][[A']] = e^{ip\cdot x}(A \cdot \psi)([A'])
\]
or, in more detail,
\[ [U(x, A)\psi](A') = e^{(p, x) \cdot \mathcal{D}(j/2)}(u([A'])^{-1}Au(A^{-1} \cdot [A']))\psi(A^{-1} \cdot [A']). \quad (2.74) \]

Finally, we would like to replace the homogeneous manifold \( \text{SL}(2, \mathbb{C})/\text{SU}(2) \) with the mass hyperboloid \( X_m^+ \). Composing with a diffeomorphism of \( X_m^+ \) onto \( \text{SL}(2, \mathbb{C})/\text{SU}(2) \) we can regard \( s \) as a section of a vector bundle over \( X_m^+ \) and \( \psi \) as a function on \( X_m^+ \). Specifically, if we let \( P_0 = (m, 0, 0, 0) \), then

\[ \beta_{P_0} : \text{SL}(2, \mathbb{C})/\text{SU}(2) \to X_m^+ \]

is a diffeomorphism.

**Exercise 2.8.3.** Show that \( A' \cdot P_0 = p \), where \( p \) is the same point of \( X_m^+ \) that appears in the exponential factor \( e^{(p, x) \cdot \mathcal{D}(j/2)} \) of \( [U(x, A)\psi](A') \).

Then

\[ \pi_m = \beta_{P_0} \circ \pi_{D(j/2)} : \text{SL}(2, \mathbb{C}) \times_{D(j/2)} \mathcal{C}(D(j/2)) \to X_m^+ \]

is a \( C^0 \)-Hilbert bundle over \( X_m^+ \),

\[ s_m = s \circ \beta^{-1}_{P_0} : X_m^+ \to \text{SL}(2, \mathbb{C}) \times_{D(j/2)} \mathcal{C}(D(j/2)) \]

is the section of this vector bundle corresponding to \( s \),

\[ \omega = u \circ \beta^{-1}_{P_0} : X_m^+ \to \text{SL}(2, \mathbb{C}) \]

is the smooth section of the principal \( \text{SU}(2) \)-bundle \( \beta_{P_0} \circ \pi : \text{SL}(2, \mathbb{C}) \to X_m^+ \) corresponding to \( u \), and

\[ \psi_m = \psi \circ \beta^{-1}_{P_0} : X_m^+ \to \mathcal{C}(D(j/2)) \]

is the function on \( X_m^+ \) corresponding to \( \psi \).

Notice that, since \( s \) is an \( L^2 \)-section with respect to the invariant measure \( \mu \) on \( \text{SL}(2, \mathbb{C})/\text{SU}(2) \) and since \( \mu \) is the pushforward measure of \( \mu_m \) on \( X_m^+ \) by the diffeomorphism \( \beta^{-1}_{P_0} \),

\[ \psi_m \in L^2(X_m^+, \mu_m, \mathcal{C}(D(j/2))). \]

Since \( \psi_m = \psi \circ \beta^{-1}_{P_0} \) we define the action of \( \text{SL}(2, \mathbb{C}) \) on \( \psi_m \) by

\[ A \cdot \psi_m = (A \cdot \psi) \circ \beta^{-1}_{P_0} \]
for every \( A \in \text{SL}(2, \mathbb{C}) \). Then \((A \cdot \psi_m)(p) = (A \cdot \psi)([A'])\), where \( p = \beta_{P_0}([A']) = A' \cdot P_0 \). Furthermore, \( u([A']) = u(\beta_{P_0}^{-1}(p)) = \omega(p) \).

**Exercise 2.8.4.** Show that \( \beta_{P_0}(A^{-1} \cdot [A']) = A^{-1} \cdot p \).

However, \( \text{SL}(2, \mathbb{C}) \) acts on \( X^+_m \) by Lorentz transformations via the double covering \( \kappa \) so, if we write \( \Lambda_A \in \mathcal{L}^1 \) for the image \( \kappa(A) \) of \( A \), then what the last Exercise shows is that

\[
A^{-1} \cdot [A'] = \beta_{P_0}^{-1}(\Lambda_A^{-1} \cdot p).
\]

Substituting all of this into (2.73) gives

\[
(A \cdot \psi_m)(p) = \mathcal{D}^{(j/2)} \left( \omega(p)^{-1} A \omega(\Lambda_A^{-1} \cdot p) \right) \psi_m(A^{-1} \cdot p).
\]

Thus, as an operator on \( \psi_m \in L^2(X^+_m, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)})) \),

\[
[U(x, A)\psi_m](p) = e^{i\ell(p,x)}(A \cdot \psi_m)(p)
= e^{i\ell(p,x)\mathcal{D}^{(j/2)}} \left( \omega(p)^{-1} A \omega(\Lambda_A^{-1} \cdot p) \right) \psi_m(A^{-1} \cdot p).
\]

This is the form in which one most often sees the representation \( U(x, A) \) written in the physics literature.

**Remark 2.8.4.** Just as a reminder and to anticipate some of the physical terminology that we will discuss more fully as we proceed let’s record the interpretation of the various objects in (2.76). \((x, A)\) is an element of the double cover \( \mathbb{R}^{1,3} \times \text{SL}(2, \mathbb{C}) \) of the Poincaré group \( \mathcal{P}^+ \) so \( x \) is interpreted as a translation on Minkowski spacetime and \( A \) is an element of \( \text{SL}(2, \mathbb{C}) \) acting on Minkowski spacetime by Lorentz transformations via the double covering \( \kappa : \text{SL}(2, \mathbb{C}) \to \mathcal{L}^1. \) \( U(x, A) \) is a unitary operator on \( L^2(X^+_m, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)})) \), where \( X^+_m \subseteq \mathbb{R}^{1,3} \) is the mass hyperboloid, \( \mu_m \) is an \( \text{SL}(2, \mathbb{C}) \)-invariant measure on \( X^+_m \), and \( \mathcal{C}(\mathcal{D}^{(j/2)}) \) is the finite-dimensional Hilbert space of carriers of the spin \( j/2 \) representation of \( \text{SU}(2) \). We will refer to \( L^2(X^+_m, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)})) \) as the 1-particle space. The elements of \( L^2(X^+_m, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)})) \) will be interpreted as wave functions for a free material particle with 4-momentum \( p \) satisfying \((p, p) = m^2 \) so that \( m \) is interpreted as the mass of the particle. These wave functions take values in \( \mathcal{C}(\mathcal{D}^{(j/2)}) \) which is \( \mathbb{C} \) only when \( j = 0 \). The significance of the additional components when \( j > 0 \) will be explained in due course (Section ??). The point \( p = A' \cdot (m, 0, 0, 0) \) varies over all of \( X^+_m \) as \( A' \) varies over \( \text{SL}(2, \mathbb{C}) \) and \( \omega(p) \) is a mapping of \( X^+_m \) to \( \text{SL}(2, \mathbb{C}) \) with the property that \( \omega(p) \cdot (m, 0, 0, 0) = p \) for each \( p \in X^+_m \). Physically, \( \omega(p) \) therefore corresponds to a Lorentz transformation from a frame in which the particle’s 4-momentum is \((p_0, p_1, p_2, p_3)\) to a frame in which the 4-momentum is \((m, 0, 0, 0)\), that is, to the rest frame of the particle. \( \Lambda_A = \kappa(A) \) and \( \Lambda_A^{-1} \cdot p \) is the contragredient action of \( \Lambda_A^{-1} \) on \( X^+_m \subseteq \mathbb{R}^{1,3} \), that is, \( \Lambda_A \cdot p = [(\Lambda_A^{-1})^\dagger]^\dagger \) \( p = \Lambda_A^\dagger p = \kappa(A)^\dagger p \). \( U \) itself is called the (Wigner) representation.
of mass $m$ and spin $j/2$. What the term “spin” has to do with the corresponding term in physics will be discussed in Section A.5.

Notice that, when $x = 0$ (no translation), (2.76) reduces to

$$[U(0, A)\psi_m](p) = \mathcal{D}^{(j/2)}(\omega(p)^{-1} A \omega(A^{-1}_A \cdot p)) \psi_m(A^{-1}_A \cdot p)$$

(2.77)

whereas, when $A = I = \text{id}_{\text{SL}(2, C)}$ (no Lorentz transformation),

$$[U(x, I)\psi_m](p) = e^{(p \cdot x)} \psi_m(p).$$

(2.78)

The simplest case of (2.76) occurs when $j = 0$ since $\mathcal{C}(\mathcal{D}^{(0)}) = \mathbb{C}$ and $\mathcal{D}^{(0)}$ is the trivial representation, that is, $\mathcal{D}^{(0)}(U) = \text{id}_C$ for every $U \in \text{SU}(2)$ (see Remark 1.2.3). In this case,

$$[U(x, A)\psi_m](p) = e^{(p \cdot x)} \psi_m(A^{-1}_A \cdot p)$$

(2.79)

for all $(x, A) \in \text{ISL}(2, \mathbb{C})$. Most of our attention later on will focus on this spin zero case. We will refer to a $\psi_m$ satisfying (2.79) as classical, relativistic scalar field on $X^*_m$, and our ultimate goal is to quantize the physical system it describes.

We make one more general observation about (2.76). The carriers of the representation $\mathcal{D}^{(j/2)}$ can be identified with $2^j$-tuples $(\xi^{A_1 \cdots A_j})_{A_1, \cdots, A_j = 1, 2}$ in $\mathbb{C}^{2^j}$ that are invariant under all permutations of $A_1 \cdots A_j$ (see Example 1.2.2). The dimension of this subspace of $\mathbb{C}^{2^j}$ is $j + 1$ so each $\psi_m$ has $j + 1$ components which we write as

$$\psi_m^{A_1 \cdots A_j}(p), \quad A_1, \ldots, A_j = 1, 2,$$

where $\psi_m^{A_1 \cdots A_j}(p)$ is invariant under all permutations of $A_1 \cdots A_j$. The effect of the ISL$(2, \mathbb{C})$-action is to yield a new set of components

$$\hat{\psi}_m^{A_1 \cdots A_j}(p) = [U(x, A)\psi_m]^{A_1 \cdots A_j}(p).$$

If we denote the element $\omega(p)^{-1} A \omega(A^{-1}_A \cdot p)$ of SU$(2)$ by $U = (U^A_B)_{A,B = 1, 2}$, then

$$\hat{\psi}_m^{A_1 \cdots A_j}(p) = e^{(p \cdot x)} U^{A_1 \cdot B_1} \cdots U^{A_j \cdot B_j} \psi_m^{B_1 \cdots B_j}(A^{-1}_A \cdot p),$$

(2.80)

where we sum over $B_1, \ldots, B_j = 1, 2$. We will refer to a $\psi_m$ with component functions that transform under the ISL$(2, \mathbb{C})$-action according to (2.80) as a (contravariant) spinor field of rank $j$ on $X^*_m$.

All of our efforts in this section have been directed toward the irreducible, unitary representation of the double cover ISL$(2, \mathbb{C})$ of the Poincaré group $\mathcal{P}^1_\mp$. We have seen in Section 2.7 how these determine the irreducible projective representations of $\mathcal{P}^1_\mp$ and that these are the objects that express relativistic invariance for quantum systems. We should note that we have, in fact, also determined the irreducible, unitary
representations of $\mathfrak{p}_+^I$ itself. Certainly, every such representation gives rise to one of the representations of ISL(2, $\mathbb{C}$) that we have found by simply composing with the covering homomorphism $\kappa : \text{ISL}(2, \mathbb{C}) \to \mathfrak{p}_+^I$. On the other hand, a representation of ISL(2, $\mathbb{C}$) will descend to a representation of $\mathfrak{p}_+^I$ precisely when $U(x, -A) = U(x, A)$ for all $(x, A) \in \text{ISL}(2, \mathbb{C})$.

**Exercise 2.8.5.** Show that a Wigner representation of ISL(2, $\mathbb{C}$) descends to a representation of $\mathfrak{p}_+^I$ if and only if it has integral spin, that is, if and only if $j$ is even.

We will conclude this section by writing out in a bit more detail a few special cases. First we will consider the effect of a pure translation by $x \in \mathbb{R}^{1,3}$ given by (2.78). The generators of translations on $\mathbb{R}^{1,3}$ are the elements $P_{\mu}, \mu = 0, 1, 2, 3$, of the Lie algebra $\mathfrak{p}$. Let’s consider, for example, the generator $P_1$ of translations in the $x^1$-direction in $\mathbb{R}^{1,3}$. Then, for any $t \in \mathbb{R}$,

$$e^{itP_1} = e^{O_1} = tO_1$$

which is the translation in $\mathbb{R}^{1,3}$ by $x = (0, t, 0, 0)$ (see (2.23)). Thus, for any $p = (p_0, p_1, p_2, p_3) \in X_1^+$,

$$e^{i(p, x)} = e^{-itP_1}$$

and so, by (2.78),

$$[U((0, t, 0, 0), I)\psi_m](p) = e^{-itP_1}\psi_m(p).$$

Thus, $U((0, t, 0, 0), I), t \in \mathbb{R}$, is a strongly continuous, 1-parameter group of unitary multiplication operators on $L^2(X_1^+, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)}))$ so, by Stone’s Theorem (Section VIII.4 of [RS1]), there is a unique self-adjoint operator $\hat{P}_1$ on $L^2(X_1^+, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)}))$ with $U((0, t, 0, 0), I) = \exp(it\hat{P}_1)$ and $\hat{P}_1$ is determined by

$$(i\hat{P}_1\psi_m)(p) = \lim_{t \to 0} \left( \frac{e^{-itP_1}\psi_m(p) - \psi_m(p)}{t} \right) = -ip_1\psi_m(p).$$

Consequently, $\hat{P}_1$ is just the multiplication operator

$$(\hat{P}_1\psi_m)(p) = -p_1\psi_m(p)$$

on $L^2(X_1^+, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)}))$. The domain of $\hat{P}_1$ is the set of all $\psi_m$ in $L^2(X_1^+, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)}))$ for which $p_1\psi_m(p)$ is also in $L^2(X_1^+, \mu_m, \mathcal{C}(\mathcal{D}^{(j/2)}))$. Similarly,

$$[U((0, 0, t, 0), I)\psi_m](p) = e^{-itP_2}\psi_m(p),$$

$$[U((0, 0, 0, t), I)\psi_m](p) = e^{-itP_3}\psi_m(p),$$

and so on.
[\[ U((t,0,0,0),I)\psi_m](p) = e^{itp_0}\psi_m(p). \]

and the corresponding self-adjoint operators \( \hat{P}_2, \hat{P}_3, \) and \( \hat{P}_0 \) are multiplication by \(-p_2, -p_3 \) and \( p_0 \), respectively (note that there is no minus sign in the exponent for time translations). All of the self-adjoint operators \( \hat{P}_{\mu}, \mu = 0, 1, 2, 3 \), share a common, invariant, dense subspace \( \mathcal{D} \) on which they are essentially self-adjoint (the smooth functions on \( X_m^+ \) with compact support). On this subspace each of the operators \( \hat{P}_{\mu}^2 \) is defined and essentially self-adjoint since it is just multiplication by \( p_\mu^2 \). Also on this subspace the operator

\[ \hat{P}_0^2 - \hat{P}_1^2 - \hat{P}_2^2 - \hat{P}_3^2 \]

is multiplication by

\[ p_0^2 - p_1^2 - p_2^2 - p_3^2 = \langle p, p \rangle \]

and this is constant and equal to \( m^2 \) on \( X_m^+ \). Writing \( \hat{P}^0 = \eta^\mu \hat{P}_\mu \) we will refer to

\[ \hat{M}^2 = \hat{P}_\mu^2 \]

as the (squared) mass operator on \( L^2(X_m^+, \mu_m, \mathcal{D}(j/2)) \). Notice that \( \hat{M}^2 \) is the operator corresponding to the first Casimir invariant of \( \mathfrak{p} \) which lives, not in \( \mathfrak{p} \), but in its universal enveloping algebra \( \mathfrak{u}(\mathfrak{p}) \). It acts by scalar multiplication and the eigenvalue is just the squared mass \( m^2 \). The skew-adjoint operators corresponding to \( \hat{P}_{\mu}, \mu = 0, 1, 2, 3 \), are

\[ dU(P_\mu) = i\hat{P}_\mu, \quad \mu = 0, 1, 2, 3 \]

so \( dU \) qualifies as a realization of the Lie subalgebra of \( \mathfrak{p} \) generated by the \( P_{\mu}, \mu = 0, 1, 2, 3 \).

One would like to extend this to a realization of \( \mathfrak{p} \) itself. For this one needs to examine the remaining generators of \( \mathfrak{p} \). The images of these under \( U \) are given by (2.77) which, for convenience, we repeat here.

\[ [U(0,A)\psi_m](p) = \mathcal{D}(j/2)(\omega(p)^{-1} A \omega(A_A^{-1} \cdot p))\psi_m(A_A^{-1} \cdot p) \quad (2.81) \]

Here \( \omega : X_m^+ \to \text{SL}(2, \mathbb{C}) \) is any function with the property that \( \omega(p) \cdot (m,0,0,0) = p \) for each \( p \in X_m^+ \). In Remark 2.6.3 we showed that one can choose an \( \omega \) that is equivariant with respect to the natural actions of \( \text{SU}(2) \) on \( X_m^+ \) and \( \text{SL}(2, \mathbb{C}) \) and we will now assume that we have made such a choice. In particular, if \( X \) is any element of the Lie algebra of the isotropy subgroup \( \text{SU}(2) \) of \( (m,0,0,0) \) and \( t \in \mathbb{R} \), then

\[ \omega(e^{itX} \cdot p) = e^{itX} \omega(p) e^{-itX}. \]

Exercise 2.8.6. Show that, with such a choice for \( \omega \), (2.81) gives
2.8 Positive Energy Representations of $\mathfrak{p}^+$ with $m > 0$

\[ [U(0, e^{iX})\psi_m](p) = D^{(j/2)}(e^{iX}) \psi_m(e^{-iX} \cdot p) \] (2.82)

for any $X \in \text{su}(2)$.

We want to compute the self-adjoint operator $\hat{A}$ corresponding to the 1-parameter group $U(0, e^{iX})$ of unitary operators on $L^2(X_m^+, \mathcal{C}(D^{(j/2)}))$. The dimension of the space $\mathcal{C}(D^{(j/2)})$ of carriers of the representation $D^{(j/2)}$ of SU(2) is $j + 1$ (Example 1.2.1) so we can regard $\psi_m(e^{-iX} \cdot p)$ as a $(j + 1)$ column vector of functions of $t$ for each $p \in X_m^+$ and $D^{(j/2)}(e^{iX})$ as a $(j + 1) \times (j + 1)$ matrix of smooth functions of $t$. If $\psi_m$ is a smooth element of $L^2(X_m^+, \mathcal{C}(D^{(j/2)}))$ and if we compute the $t$-derivatives of $\psi_m(e^{-iX} \cdot p)$ and $D^{(j/2)}(e^{iX})$ componentwise and entrywise, respectively, then

\[
[i\hat{A}\psi_m](p) = \lim_{t \to 0} \frac{[U(0, e^{iX})\psi_m](p) - \psi_m(p)}{t}
\]

\[
= \frac{d}{dt} [D^{(j/2)}(e^{iX})\psi_m(e^{-iX} \cdot p)]|_{t=0}
\]

\[
= \frac{d}{dt} [D^{(j/2)}(I)\psi_m(e^{-iX} \cdot p)]|_{t=0} + \frac{d}{dt} [D^{(j/2)}(e^{iX})]\psi_m(p)
\]

\[
= \frac{d}{dt} \psi_m(e^{-iX} \cdot p)|_{t=0} + \frac{d}{dt} [D^{(j/2)}(e^{iX})]\psi_m(p)
\]

so that $\hat{A}$ splits into the sum of two operators

\[
-i \frac{d}{dt} \psi_m(e^{-iX} \cdot p)|_{t=0} - i \frac{d}{dt} [D^{(j/2)}(e^{iX})]|_{t=0} \psi_m(p).
\] (2.83)
Exercise 2.8.7. Let $\kappa : \text{SL}(2, \mathbb{C}) \to \mathcal{L}_+^1$ be the covering map. Prove each of the following.

\[ e^{-itM_{23}} = e^{iM_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos t & -\sin t \\ 0 & 0 & \sin t & \cos t \end{pmatrix} = \kappa(e^{i(t/2)\sigma_1}) \]

\[ e^{-itM_{31}} = e^{iM_2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos t & \sin t & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\sin t & \cos t & 0 \end{pmatrix} = \kappa(e^{i(t/2)\sigma_2}) \]

\[ e^{-itM_{12}} = e^{iM_3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos t & -\sin t & 0 \\ 0 & \sin t & \cos t & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \kappa(e^{i(t/2)\sigma_3}) \]

Hint: Exercise 2.4.5

Consider, for example, the (complex) generator $M_{12} = J_3 = iM_3$ of rotations about the $x^3$-axis ($M_{23}$ and $M_{31}$ are entirely analogous). We will denote the self-adjoint operator corresponding to $U(0, e^{-itM_{12}})$ by $\hat{M}_{12}$ and will denote the two operators in (2.83) by $\hat{O}_{12}$ and $\hat{S}_{12}$, that is, $\hat{M}_{12} = \hat{O}_{12} + \hat{S}_{12}$, where

\[ [\hat{O}_{12}\psi_m](p) = -i\frac{d}{dt}\psi_m(e^{iM_{12}} \cdot p)|_{t=0} \tag{2.84} \]

and

\[ [\hat{S}_{12}\psi_m](p) = -i\frac{d}{dt}[D^{(j/2)}(e^{-iM_{12}})]|_{t=0}\psi_m(p). \tag{2.85} \]

Thus,

\[ [\hat{O}_{12}\psi_m](p) = -i\frac{d}{dt}\left[\psi_m(p_0, p_1 \cos t - p_2 \sin t, p_1 \sin t + p_2 \cos t, p_3)\right]|_{t=0} \]

\[ = -i\left(p_2 \frac{\partial \psi_m}{\partial p_1} - p_1 \frac{\partial \psi_m}{\partial p_2}\right) \]

so $\hat{O}_{12}$ is the unique self-adjoint extension of

\[ -i\left(p_2 \frac{\partial}{\partial p_1} - p_1 \frac{\partial}{\partial p_2}\right) \]

on $L^2(\mathbb{X}_m^+, \mu, \mathcal{C}(D^{(j/2)}))$. We will refer to the physical quantity represented by $\hat{O}_{12}$ as the orbital angular momentum about the $x^3$-axis corresponding to the representation $U$ and the section $\omega$. 
The operator $\hat{S}_{12}$ depends, of course, on the spin $s = \frac{j}{2} \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \ldots\}$ of the representation. When $s = 0$, $\mathcal{C}(\mathcal{D}^{(0/2)}) = \mathbb{C}$ and $\mathcal{D}^{(0/2)}$ is the trivial representation of $SU(2)$ that sends every $U \in SU(2)$ to $id_C$ so $\hat{S}_{12}$ is identically zero.

$$-i \frac{d}{dt} \left[ \mathcal{D}^{(0/2)}(e^{-itM_{12}}) \right] \bigg|_{t=0} = 0 \quad (s = 0)$$

When $s = \frac{1}{2}$, $\mathcal{C}(\mathcal{D}^{(1/2)}) = \mathbb{C}^2$ and $\mathcal{D}^{(1/2)}$ is the standard representation of $SU(2)$ that sends every $U \in SU(2)$ to $U$ acting on $\mathbb{C}^2$ by matrix multiplication. If we identify $M_3$ with the basis element $X_3$ of the Lie algebra $so(3)$ and this, in turn, with the basis element $-\frac{i}{2}\sigma_3$ of $su(2)$ (see Remark 2.5.1), then

$$-i \frac{d}{dt} \left[ \mathcal{D}^{(1/2)}(e^{-itM_{12}}) \right] \bigg|_{t=0} = -i \frac{d}{dt} \begin{pmatrix} e^{-it/2} & 0 \\ 0 & e^{it/2} \end{pmatrix} \bigg|_{t=0} = \begin{pmatrix} -\frac{i}{2} & 0 \\ 0 & \frac{i}{2} \end{pmatrix} \quad (s = \frac{1}{2})$$

Notice that the eigenvalues of this matrix are $\{-\frac{1}{2}, \frac{1}{2}\}$. The action of $\hat{S}_{12}$ on $\psi_m = \begin{pmatrix} \psi^1_m \\ \psi^2_m \end{pmatrix}$ is therefore as follows.

$$[\hat{S}_{12}\psi_m](p) = \begin{pmatrix} -\frac{i}{2} & 0 \\ 0 & \frac{i}{2} \end{pmatrix} \begin{pmatrix} \psi^1_m(p) \\ \psi^2_m(p) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -\psi^1_m(p) \\ \psi^2_m(p) \end{pmatrix} \quad (s = \frac{1}{2})$$

**Exercise 2.8.8.** Show that, when $s = 1$,

$$-i \frac{d}{dt} \left[ \mathcal{D}^{(j/2)}(e^{-itM_{12}}) \right] \bigg|_{t=0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (s = 1)$$

so that the eigenvalues are $\{-1, 0, 1\}$. *Hint:* See Exercise 1.2.8.

Continuing in this way one finds that if $s = \frac{j}{2}$, then $-i \frac{d}{dt} \left[ \mathcal{D}^{(j/2)}(e^{-itM_{12}}) \right] \bigg|_{t=0}$ has eigenvalues

$$\left\{ -\frac{j}{2}, -\frac{j}{2} + 1, \ldots, -\frac{j}{2} + j = \frac{j}{2} \right\}$$

so that the spin of the representation is the largest eigenvalue of $\hat{S}_{12}$. We will refer to $\hat{S}_{12}$ as the spin angular momentum about the $x^3$-axis corresponding to the representation $U$ and the section $\omega$. $M_{12} = \hat{O}_{12} + \hat{S}_{12}$ is the total angular momentum about the $x^3$-axis.
Remark 2.8.5. An element $\psi_m$ of the 1-particle space $L^2(X^s, \mu, C(D^{(j/2)}))$ is interpreted as the wave function of a free material particle of “mass” $m$ and “spin” $s = \frac{j}{2}$, where “mass” and “spin” here refer to the terms as they are used in physics to denote certain physically measured observables. One would like to understand the rationale behind identifying these physical observables with the mathematical mass and spin parameters that we have introduced here. In the case of the mass we have seen above that this rationale is essentially just the relativistic relation between the Minkowski norm of a particle’s 4-momentum and its mass.

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2$$

The corresponding rationale for spin will, of course, have to wait until we know what the physicists mean by “spin” and so must be deferred to Section A.5. One particular aspect of this rationale is worth briefly anticipating here however. We have identified the spin of a representation with the largest eigenvalue of the spin angular momentum about the $x^3$-axis and one might reasonably wonder what is so special about the $x^3$-axis. The answer is quite simple, that is, nothing at all. Physically, the “spin” of, say, an electron about the $x^3$-axis is thought of as the third component of a “spin vector”, but this is a rather peculiar “vector” that could only exist in the quantum world in that its projection onto every axis is either $\frac{1}{2}$ or $-\frac{1}{2}$ ($\pm \frac{1}{2}$ if $\hbar$ is not taken to be 1). One can see that this is at least consistent with the mathematical notion of spin that we have introduced. For example, if one repeats the arguments we have just given in the $s = \frac{1}{2}$ case with $M_{12}$ replaced by the generator $M_{23}$ of rotations about the $x^1$-axis one finds that

$$-i\frac{d}{dt} \left[ D^{(j/2)}(e^{-itM_{23}}) \right] \bigg|_{t=0} = -i\frac{d}{dt} \left( \begin{array}{cc} \cos(t/2) & -i \sin(t/2) \\ -i \sin(t/2) & \cos(t/2) \end{array} \right) \bigg|_{t=0} = \left( \begin{array}{cc} 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{array} \right) \ (s = \frac{1}{2})$$

(Exercise 2.4.5) and this has eigenvalues $\pm \frac{1}{2}$. Consequently, $\hat{S}_{12}$ and $\hat{S}_{23}$ have the same eigenvalues and, in particular, the same largest eigenvalue. The spin about the $x^1$-axis is the same as the spin about the $x^3$-axis.

Exercise 2.8.9. Prove the same thing for $\hat{S}_{31}$.

One can generalize to show that, when the representation of SU(2) is $D^{(j/2)}$, then $s = \frac{j}{2}$ is the largest eigenvalue of all of the operators $\hat{S}_{12}$, $\hat{S}_{23}$, and $\hat{S}_{31}$ so that, in this sense at least, our spin parameter has the property required of the physicist’s “spin” observable.
Appendix A
Physical Background

A.1 Introduction

Although this volume was planned as a sequel to [Nab5] we would not want this earlier work to be a *sine qua non* for this one. Toward this end we will try to remove some potential obstacles by providing brief synopses of a number of topics in classical and quantum mechanics that we must depend upon here and that are discussed in greater detail in [Nab5].

A.2 Finite-Dimensional Lagrangian Mechanics

Section 2.2 of [Nab5] describes the Lagrangian formulation of classical particle mechanics in some detail with numerous examples. Here we will simply collect together a summary of those items we need in order to pursue our current objectives and conclude with one example that, we hope, will solidify the ideas.

One begins with an $n$-dimensional smooth manifold $M$ called the *configuration space* and generally denotes a local coordinate system on $M$ by $(q^1, \ldots, q^n)$. We think of $M$ as the space of possible positions of the particles in the system. For two particles moving in $\mathbb{R}^3$, for example, $M = \mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$. The tangent bundle $TM$ of $M$ (Remark 2.2.5 of [Nab5] or Section 1.25 of [Warn]) is called the *state space* and the points $(x, v_x)$ in it represent possible *states* of the system with $x \in M$ representing a possible configuration of the particles in the system and $v_x \in T_x(M)$ a possible rate of change of the configuration at $x$. The local coordinate functions $q^1, \ldots, q^n$ on $M$ together with their coordinate velocity vector fields $\partial_{q^1}, \ldots, \partial_{q^n}$ determine *natural coordinates* $(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n)$ on $TM$. These are defined in the following way. For each $(x, v_x) \in TM$, define

$$q^i(x, v_x) = q^i(x), \quad i = 1, \ldots, n.$$

and then write

$$v_x = v_x[q^1]\partial_{q^1}(x) + \cdots + v_x[q^n]\partial_{q^n}(x) = v_x[q^i]\partial_{q^i}(x)$$

and define
\[ \dot{q}'(x, v_x) = v_x[q'] \quad i = 1, \ldots, n. \]

We will, on occasion, write \( \dot{q}'(x, v_x) \) simply as \( \dot{q}'(v_x) \).

Any smooth real-valued function \( L : TM \to \mathbb{R} \) on the state space \( TM \) is called a Lagrangian on \( M \). Such a function can be described locally in natural coordinates. We adopt the usual custom of writing such a local coordinate representation as

\[ L(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n) = L(q, \dot{q}). \]

For \( t_0 < t_1 \) in \( \mathbb{R} \) and \( a, b \in M \) the path space \( C^\infty_{a,b}([t_0, t_1], M) \) is the space of all smooth curves \( \alpha : [t_0, t_1] \to M \) with \( \alpha(t_0) = a \) and \( \alpha(t_1) = b \). Every \( \alpha \) in \( C^\infty_{a,b}([t_0, t_1], M) \) has a unique lift to a smooth curve

\[ \tilde{\alpha} : [t_0, t_1] \to TM \]

in the tangent bundle defined by

\[ \tilde{\alpha}(t) = (\alpha(t), \dot{\alpha}(t)), \]

where \( \dot{\alpha}(t) \) denotes the velocity (tangent) vector to \( \alpha \) at \( t \). The action functional associated with the Lagrangian \( L \) is the real-valued function

\[ S_L : C^\infty_{a,b}([t_0, t_1], M) \to \mathbb{R} \]

defined by

\[ S_L(\alpha) = \int_{t_0}^{t_1} L(\tilde{\alpha}(t)) \, dt = \int_{t_0}^{t_1} L(\alpha(t), \dot{\alpha}(t)) \, dt. \]

For any \( \alpha \in C^\infty_{a,b}([t_0, t_1], M) \) we define a (fixed endpoint) variation of \( \alpha \) to be a smooth map

\[ \Gamma : [t_0, t_1] \times (-\epsilon, \epsilon) \to M \]

for some \( \epsilon > 0 \) such that

\[ \Gamma(t, 0) = \alpha(t), \quad t_0 \leq t \leq t_1 \]
\[ \Gamma(t_0, s) = \alpha(t_0) = a, \quad -\epsilon < s < \epsilon \]
\[ \Gamma(t_1, s) = \alpha(t_1) = b, \quad -\epsilon < s < \epsilon. \]

For any fixed \( s \in (-\epsilon, \epsilon) \) the map

\[ \gamma_s : [t_0, t_1] \to M \]

defined by

\[ \gamma_s(t) = \Gamma(t, s) \]
is an element of $C^\infty_a([t_0, t_1], M)$. Then $S_L(\gamma_s)$ is a smooth real-valued function of the real variable $s$ whose value at $s = 0$ is $S_L(\alpha)$. We say that $\alpha \in C^\infty_a([t_0, t_1], M)$ is a stationary point, or critical point of the action functional $S_L$ if

$$\frac{d}{ds} S_L(\gamma_s) \bigg|_{s=0} = 0$$

for every variation $\Gamma$ of $\alpha$. In this case we will call $\alpha$ a stationary curve, or a critical curve for the action functional $S_L$. According to the Principle of Stationary Action, also called the Principle of Least Action, the actual time evolution of the state of the system will take place along the lift of a stationary curve for $S_L$. One should understand, of course, that this is not a mathematical theorem, but rather a physical assumption that happens to be well born out by experience. We will offer some motivation in Example A.2.1.

For curves $\alpha$ that lie in some coordinate neighborhood in $M$ one can write down explicit equations that are necessary conditions for $\alpha$ to be a stationary point of $S_L$. Let $\alpha \in C^\infty_a([t_0, t_1], M)$ be a curve whose image lies in a coordinate neighborhood with coordinate functions $q^1, \ldots, q^n$. The lift $\tilde{\alpha}$ of $\alpha$ is written in natural coordinates as $\tilde{\alpha}(t) = (q^1(t), \ldots, q^n(t), \dot{q}^1(t), \ldots, \dot{q}^n(t))$, where $q^i(t)$ is a notational shorthand for $q^i(\alpha(t))$ and similarly $\dot{q}^i(t)$ means $\dot{q}^i(\dot{\alpha}(t))$. Then it is shown in Section 2.2 of [Nab5] that if $\alpha$ is a stationary point of $S_L$, then

$$\frac{\partial L}{\partial q^i}(\alpha(t), \dot{\alpha}(t)) - \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}^i}(\alpha(t), \dot{\alpha}(t)) \right) = 0, \quad 1 \leq i \leq n. \quad (A.1)$$

These are the famous Euler-Lagrange equations which one often sees written simply as

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}^i} \right) = 0, \quad 1 \leq i \leq n. \quad (A.2)$$

The Euler-Lagrange equations are satisfied along any stationary curve for $S_L$.

Notice that one can draw a rather remarkable conclusion just from the form in which the Euler-Lagrange equations are written, namely, that if the Lagrangian $L$ happens not to depend on one of the coordinates, say $q^j$, then $\frac{\partial L}{\partial q^j} = 0$ everywhere and (A.1) implies that, along any stationary curve, $\frac{\partial L}{\partial \dot{q}^j}$ is constant. In more colloquial terms, $\frac{\partial L}{\partial \dot{q}^j}$ is conserved as the system evolves.

$$\frac{\partial L}{\partial \dot{q}^j} = 0 \implies \frac{\partial L}{\partial \dot{q}^j} \text{ is conserved along any stationary curve.}$$

In the physics literature

$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

is called the momentum conjugate to $q^i$. 

A number of concrete examples of such conservation laws in classical mechanics are written out in Section 2.2 of [Nab5] and we will see a few of these as we proceed. We have here the simplest instance of one of the most important features of the Lagrangian formalism, that is, the deep connection between the symmetries of a Lagrangian (in this case, its invariance under translation of $q^i$) and the existence of quantities that are conserved during the evolution of the system. The next order of business is to describe a more general result giving rise to such conserved quantities.

If $L : TM \to \mathbb{R}$ is a Lagrangian on a smooth manifold $M$, then a symmetry of $L$ is a diffeomorphism

$$F : M \to M$$

of $M$ onto itself for which the induced map

$$TF : TM \to TM$$

given by

$$TF(x, v_x) = (F(x), F_{x\cdot}(v_x)),$$

where $F_{x\cdot}$ is the derivative of $F$ at $x$, satisfies

$$L \circ TF = L.$$

A symmetry (or, rather, its induced map on the state space) carries one state of the system onto another state at which the value of the Lagrangian is the same.

An “infinitesimal symmetry” of $L$ is essentially a 1-parameter family of symmetries arising from the 1-parameter group of diffeomorphisms of a smooth vector field on $M$ (Section 3.5 of [BG] or Section 5.7 of [Nab2]). The precise definition is complicated just a bit by the fact that not every vector field on $M$ is complete, that is, has integral curves defined for all $t \in \mathbb{R}$. In order not to cloud the essential issues we will give the definition twice, once for vector fields that are complete and once for those that need not be complete (naturally, the first definition is a special case of the second).

Let $L$ be a Lagrangian on a smooth manifold $M$. A complete vector field $X$ on $M$ is said to be an infinitesimal symmetry of $L$ if each $\varphi_t$ in the 1-parameter group of diffeomorphisms of $X$ is a symmetry of $L$. Now we drop the assumption that $X$ is complete. For each $x \in M$, let $\alpha_{x\cdot}$ be the maximal integral curve of $X$ through $x$ (Theorem 3.4.1 of [BG] or Theorem 5.7.2 of [Nab2]). For each $t \in \mathbb{R}$, let $D_t$ be the set of all $x \in M$ for which $\alpha_{x\cdot}$ is defined at $t$ and define $\varphi_t : D_t \to M$ by $\varphi_t(x) = \alpha_{x\cdot}(t)$. By Theorem 5.7.4 of [Nab2], each $D_t$ is an open set (perhaps empty) and, if $D_t \neq \emptyset$, then $\varphi_t$ is a diffeomorphism of $D_t$ onto $D_{-t}$ with inverse $\varphi_{-t}$. Now we will say that $X$ is an infinitesimal symmetry of $L$ if, for each $t$ with $D_t \neq \emptyset$, the induced map $T\varphi_t : T D_t \to T D_{-t}$, defined by $(T\varphi_t)(x, v_x) = (\varphi_t(x), (\varphi_t)_\cdot(v_x))$, satisfies $L \circ T\varphi_t = L$ on $T D_t$. 
The content of our next result is that infinitesimal symmetries of $L$ give rise to conserved quantities.

**Theorem A.2.1. (Noether’s Theorem)** Let $L : TM \to \mathbb{R}$ be a Lagrangian on a smooth manifold $M$ and suppose $X$ is an infinitesimal symmetry of $L$. Let $q^1, \ldots, q^n$ be any local coordinate system for $M$ with corresponding natural coordinates $\dot{q}^1, \ldots, \dot{q}^n$. Write $X = X^1 \partial_{q^1} + \cdots + X^n \partial_{q^n} = X^i \partial_{\dot{q}^i}$. Then

$$X^i \frac{\partial L}{\partial \dot{q}^i} = X^i p_i = X^1 p_1 + \cdots + X^n p_n$$

is constant along every stationary curve in the coordinate neighborhood on which $q^1, \ldots, q^n$ are defined.

Notice that if we have a Lagrangian $L$ that is independent of one of the coordinates in $M$, say, $q_i$, then certainly $X = \partial_{q_i}$ is an infinitesimal symmetry. Since the only component of $X$ relative to $\partial_{q^1}, \ldots, \partial_{q^n}$ is the $i$th and this is 1 we find that the corresponding Noether conserved quantity is the same as the one we found earlier, namely, the conjugate momentum $p_i = \frac{\partial L}{\partial \dot{q}^i}$.

Infinitesimal symmetries often arise in the following way. Recall that a left action of a Lie group $G$ on $M$ is a smooth map $\sigma : G \times M \to M$, usually written $(g \cdot x) = e^{-tN} \cdot x$ for all $g_1, g_2 \in G$ and all $x \in M$. Given such an action one can define, for each $g \in G$, a diereomorphism $\sigma_g : M \to M$ by $\sigma_g(x) = g \cdot x$. If a Lagrangian is given on $M$, then it may be possible to find a Lie group $G$ and a left action $\sigma$ of $G$ on $M$ for which these diffeomorphisms $\sigma_g$ are all symmetries of $L$. In this case we refer to $G$ as a symmetry group of $L$. If $\mathfrak{g}$ is the Lie algebra of $G$, then each nonzero element $N$ of $\mathfrak{g}$ gives rise to an infinitesimal symmetry $X_N$ defined at each $x \in M$ by

$$X_N(x) = \left. \frac{d}{dt} \sigma(tN \cdot x) \right|_{t=0}.$$

Each of these in turn gives rise, via Noether’s Theorem, to a conserved quantity. The conservation laws come from the Lie algebra of the symmetry group.

We will now try to illustrate all of these ideas with a concrete example. Many more such examples are available in Section 2.2 of [Nab5].

**Example A.2.1. (Momentum, Angular Momentum, and Energy)** For our configuration space we begin with $M = \mathbb{R}^n$ and choose global standard Cartesian coordinates $(q^1, \ldots, q^n)$ on $\mathbb{R}^n$. The state space is then $T \mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ and the corresponding natural coordinates are $\dot{q}^1, \ldots, \dot{q}^n$. Letting $V(q^1, \ldots, q^n)$ denote an arbitrary smooth, real-valued function on $\mathbb{R}^n$ and $m$ a positive constant, we take our Lagrangian to be
\[
L(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n) = \frac{1}{2} m \sum_{i=1}^{n} (\dot{q}^i)^2 - V(q^1, \ldots, q^n).
\]

When evaluated on the lift of some \( \alpha \in C^\infty_{\omega}(I, \mathbb{R}^n) \) this gives the kinetic minus the potential energy of a particle of mass \( m \) moving along \( \alpha \) as a function of \( t \). The Principle of Stationary Action dictates that the actual trajectory of the particle is a stationary curve for the action \( S_L \). To write down the Euler-Lagrange equations we note that

\[
\frac{\partial L}{\partial q^i} = -\frac{\partial V}{\partial q^i}, \quad i = 1, \ldots, n
\]

and

\[
\frac{\partial L}{\partial \dot{q}^i} = m \dot{q}^i, \quad i = 1, \ldots, n.
\]

Thus, (A.2) becomes

\[
-\frac{\partial V}{\partial q^i} - \frac{d}{dt}(m\dot{q}^i) = 0, \quad i = 1, \ldots, n,
\]

that is,

\[
m \frac{d^2 q^i}{dt^2} = -\frac{\partial V}{\partial q^i}, \quad i = 1, \ldots, n
\]

and these are just the components of Newton's Second Law for a conservative force \(-\partial V/\partial q^i, i = 1, \ldots, n\), with potential \( V \). This is, of course, one of the primary motivations behind the Principle of Stationary Action.

Notice that if the potential \( V(q^1, \ldots, q^n) \) happens not to depend on, say, the \( i^{th} \) coordinate \( q^i \), then, since \( p_i = \frac{d}{dt} = m\dot{q}^i \), we conclude that \( p_i = m\dot{q}^i \) is constant along the trajectory of the particle. Now, \( m\dot{q}^i \) is what physicists call the \( i^{th} \) component of the particle's (linear) momentum. Thus, if the potential \( V \) is independent of the \( i^{th} \)-coordinate, then the \( i^{th} \)-component of momentum is conserved during the motion. The conjugate momentum is really momentum in this case. In particular, for a particle that is not subject to any forces (a free particle), the potential \( V(q^1, \ldots, q^n) \) is constant so all of the momentum components are conserved. One can phrase this in the following way. If the potential \( V(q^1, \ldots, q^n) \) and therefore the Lagrangian \( L(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n) = \frac{1}{2} m \sum_{i=1}^{n} (\dot{q}^i)^2 - V(q^1, \ldots, q^n) \) is invariant under translations in \( \mathbb{R}^n \), then momentum is conserved.

**Spatial Translation Symmetry implies Conservation of (Linear) Momentum**

We conclude by looking at a somewhat less obvious example of this sort of conservation law.

We will specialize to the case in which \( n = 3 \) and the potential is _spherically symmetric_, that is, \( V \) depends only on \( ||q|| = (q^1)^2 + (q^2)^2 + (q^3)^2 \)
\[ V(q^1, q^2, q^3) = V(||q||) \]

Thus, we can write the Lagrangian as

\[ L(q, \dot{q}) = \frac{1}{2}m||\dot{q}||^2 - V(||q||). \]

Now we will find a symmetry group of \( L \). Recall that the rotation group \( \text{SO}(3) \) consists of all \( 3 \times 3 \) real matrices \( A \) that are orthogonal (\( A^T A = AA^T = \text{id}_{3 \times 3} \)) and have determinant one (\( \det A = 1 \)). \( \text{SO}(3) \) acts on \( \mathbb{R}^3 \) by matrix multiplication

\[ \sigma_A(q) = Aq, \]

where \( Aq \) means matrix multiplication with \( q \in \mathbb{R}^3 \) thought of as a column vector. Since \( A \) is invertible, \( \sigma_A \) is a diffeomorphism of \( \mathbb{R}^3 \) onto \( \mathbb{R}^3 \). Moreover, since \( \sigma_A \) is linear, its derivative at each point is the same linear map (multiplication by \( A \)) once the tangent spaces are canonically identified with \( \mathbb{R}^3 \). Thus, the induced map on state space

\[ T\sigma_A : T\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow T\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3 \]

is given by

\[ (T\sigma_A)(q, \dot{q}) = (\sigma_A(q), (\sigma_A)_{\dot{q}}(\dot{q})) = (Aq, A\dot{q}). \]

This is again a diffeomorphism of \( T\mathbb{R}^3 \) onto \( T\mathbb{R}^3 \). Moreover, since \( A \) is orthogonal, \( ||Aq|| = ||q|| \) and \( ||A\dot{q}|| = ||\dot{q}|| \) so

\[ L \circ T\sigma_A = L \]

and therefore \( \text{SO}(3) \) is indeed a symmetry group of \( L \). One says simply that \( L \) is invariant under rotation.

From this symmetry group we would now like to build infinitesimal symmetries of \( L \) and compute some Noether conserved quantities. For this we need information about the Lie algebra \( \mathfrak{so}(3) \) of \( \text{SO}(3) \) and its exponential map. Now, \( \mathfrak{so}(3) \) consists of the set of all \( 3 \times 3 \), skew-symmetric, real matrices with entrywise linear operations and matrix commutator as bracket (Section 5.8 of [Nab2]). The following is Theorem 2.2.2 of [Nab5], but the proof is on pages 393-395 of [Nab2].

**Theorem A.2.2.** Let \( N \) be an element of \( \mathfrak{so}(3) \). Then the matrix exponential \( e^{tN} \) is in \( \text{SO}(3) \) for every \( t \in \mathbb{R} \). Conversely, if \( A \) is any element of \( \text{SO}(3) \), then there is a unique \( t \in [0, \pi] \) and a unit vector \( \hat{n} = (n^1, n^2, n^3) \) in \( \mathbb{R}^3 \) for which

\[ A = e^{tN} = \text{id}_{3 \times 3} + (\sin t)N + (1 - \cos t)N^2, \]

where \( N \) is the element of \( \mathfrak{so}(3) \) given by
\[ N = \begin{pmatrix} 0 & -n^3 & n^2 \\ n^3 & 0 & -n^1 \\ -n^2 & n^1 & 0 \end{pmatrix}. \]

Geometrically, one thinks of \( A = e^{tN} \) as the rotation of \( \mathbb{R}^3 \) through \( t \) radians about an axis along \( \hat{n} \) in a sense determined by the right-hand rule from the direction of \( \hat{n} \).

Now fix an \( \hat{n} \) and the corresponding \( N \) in so(3). For any \( q \in \mathbb{R}^3 \),

\[ t \to e^{tN}q \]

is a curve in \( \mathbb{R}^3 \) passing through \( q \) at \( t = 0 \) with velocity vector

\[ \frac{d}{dt}(e^{tN}q) \bigg|_{t=0} = Nq. \]

Doing this for each \( q \in \mathbb{R}^3 \) gives a smooth vector field \( X_N \) on \( \mathbb{R}^3 \) defined by

\[ X_N(q) = \frac{d}{dt}(e^{tN}q) \bigg|_{t=0} = Nq. \]

Like any (complete) vector field on \( \mathbb{R}^3 \), \( X_N \) determines a 1-parameter group of diffeomorphisms

\[ \varphi_t : \mathbb{R}^3 \to \mathbb{R}^3, \quad -\infty < t < \infty, \]

where \( \varphi_t \) pushes each point of \( \mathbb{R}^3 \) \( t \) units along the integral curve of \( X_N \) that starts there. This 1-parameter group of diffeomorphisms is also called the flow of the vector field. In the case at hand,

\[ \varphi_t(q) = e^{tN}q. \]

Notice that each \( \varphi_t \), being multiplication by some element of SO(3), is a symmetry of \( L \) so \( X_N \) is indeed an infinitesimal symmetry of \( L \).

Next we will write out a few of these vector fields explicitly. Choose, for example, \( \hat{n} = (0, 0, 1) \in \mathbb{R}^3 \). Then

\[ N = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \]

so

\[ X_N(q) = Nq = \begin{pmatrix} -q^3 \\ q^1 \\ 0 \end{pmatrix}. \]

This gives the components of \( X_N \) relative to \( \{\partial_{q^1}, \partial_{q^2}, \partial_{q^3}\} \) so the vector field \( X_N \) is just

\[ X_{12} = q^1 \partial_{q^2} - q^2 \partial_{q^1} \]  
(A.3)
Taking $\hat{n}$ to be $(1, 0, 0)$ and $(0, 1, 0)$ one obtains, in the same way, the vector fields

$$X_{23} = q^2 \partial_{q^3} - q^3 \partial_{q^2}$$  \hspace{1cm} (A.4)

and

$$X_{31} = q^3 \partial_{q^1} - q^1 \partial_{q^3}.$$  \hspace{1cm} (A.5)

Exercise A.2.1. In the physics literature the matrices $N$ corresponding to $\hat{n} = (1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$ are denoted $L_x, L_y$ and $L_z$, respectively. Show that these form a basis for $\mathfrak{so}(3)$ and satisfy the commutation relations

$$[L_x, L_y] = L_z, \quad [L_z, L_x] = L_y, \quad [L_y, L_z] = L_x,$$

where $[,]$ denotes the matrix commutator.

To find the Noether conserved quantity corresponding to the infinitesimal symmetry $X_{12}$ in standard coordinates we compute

$$X_{12}^k \frac{\partial L}{\partial \dot{q}^k} = X_{12}^1 \frac{\partial L}{\partial \dot{q}^1} + X_{12}^2 \frac{\partial L}{\partial \dot{q}^2} + X_{12}^3 \frac{\partial L}{\partial \dot{q}^3} = -q^2(\dot{q}^1) + q^1(\dot{q}^2) = m(q^1 \dot{q}^2 - q^2 \dot{q}^1).$$

For $X_{23}$ and $X_{31}$ one obtains $m(q^2 \dot{q}^3 - q^3 \dot{q}^2)$ and $m(q^3 \dot{q}^1 - q^1 \dot{q}^3)$, respectively. Thus, along any stationary curve,

$$m \left[ q^1(t)\dot{q}^2(t) - q^2(t)\dot{q}^1(t) \right]$$

$$m \left[ q^2(t)\dot{q}^3(t) - q^3(t)\dot{q}^2(t) \right]$$

$$m \left[ q^3(t)\dot{q}^1(t) - q^1(t)\dot{q}^3(t) \right]$$

are all constant. Notice that these are precisely the components of the cross product

$$\mathbf{r}(t) \times [m \mathbf{r}(t)]$$

of the position and momentum vectors of the particle and this is what physicists call its (orbital) angular momentum (with respect to the origin). Thus, angular momentum is constant for motion in a spherically symmetric potential in $\mathbb{R}^3$.

Rotational Symmetry implies Conservation of Angular Momentum

Notice also that the constancy of the angular momentum vector along the trajectory of the particle implies that the motion takes place entirely in a 2-dimensional plane in $\mathbb{R}^3$, namely, the plane with this normal vector.

We will conclude with an example of a slightly different sort. We have defined a Lagrangian to be a function on the tangent bundle $TM$, but it is sometimes convenient to allow it to depend explicitly on $t$ as well, that is, to define a Lagrangian to
be a smooth map $L : \mathbb{R} \times TM \to \mathbb{R}$. Then, for any path in the domain of a coordinate neighborhood on $M$, one would write $L = L(t, \alpha(t), \dot{\alpha}(t)), t_0 \leq t \leq t_1$, in natural coordinates. The action associated with this path is defined, as before, to be the integral of $L(t, \alpha(t), \dot{\alpha}(t))$ over $[t_0, t_1]$. Stationary points for the action are also defined in precisely the same way and one can check that the additional $t$-dependence has no effect at all on the form of the Euler-Lagrange equations, that is, stationary curves satisfy

$$\frac{\partial L}{\partial q^k}(t, \alpha(t), \dot{\alpha}(t)) - \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}^k}(t, \alpha(t), \dot{\alpha}(t)) \right) = 0, \quad 1 \leq k \leq n. \quad (A.6)$$

Write the stationary curve in local coordinates as $\alpha(t) = (q^1(t), \ldots, q^n(t))$ and the Lagrangian evaluated on the lift of this curve as

$L(t, q^1(t), \ldots, q^n(t), \dot{q}^1(t), \ldots, \dot{q}^n(t))$.

Computing $\frac{d}{dt}$ from the chain rule and (A.6) gives

$$\frac{d}{dt} (L - p_i \dot{q}^i) = \frac{\partial L}{\partial t}$$

along the stationary curve. If it so happens that $L$ does not depend explicitly on $t$, then $\frac{\partial L}{\partial t} = 0$ so $E_L = p_i \dot{q}^i - L$ is conserved along the stationary curve. Notice that for the particle Lagrangian $L(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n) = \frac{1}{2}m \sum_{i=1}^n (\dot{q}^i)^2 - V(q^1, \ldots, q^n)$

$E_L = p_i \dot{q}^i - L = \frac{1}{2}m \sum_{i=1}^n (\dot{q}^i)^2 + V(q^1, \ldots, q^n) \quad (A.7)$

which is the total energy (kinetic plus potential). We will phrase this in the following way.

**Time Translation Symmetry implies Conservation of Energy**

Here then is a synopsis of what we have concluded about symmetries and conservation laws for classical mechanical systems.

<table>
<thead>
<tr>
<th>Symmetry</th>
<th>Conservation Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Translation</td>
<td>Linear Momentum</td>
</tr>
<tr>
<td>Rotation</td>
<td>Angular Momentum</td>
</tr>
<tr>
<td>Time Translation</td>
<td>Total Energy</td>
</tr>
</tbody>
</table>
A.3 Finite-Dimensional Hamiltonian Mechanics

In Section 2.3 of [Nab5] the Hamiltonian picture of classical mechanics evolved out of the Lagrangian picture, but it is only the Hamiltonian view itself that is significant for us at the moment and not the particular route by which one arrives at it. Except for the occasional comment, such as those in Remarks A.3.1 and A.3.3 below, we will take no heed of its Lagrangian origins here.

We will start with the simplest case and then generalize. One begins, as in the Lagrangian case, with an $n$-dimensional smooth manifold $M$ called the **configuration space** and generally denotes a local coordinate system on $M$ by $(q^1, \ldots, q^n)$. We think of $M$ as the space of possible positions of the particles in the system. For example, it is shown in Section 2.2 of [Nab5] that the configuration space of a rigid body restrained to pivot about some fixed point is the manifold $SO(3)$ and any such motion of the rigid body is represented by a continuous curve $t \in \mathbb{R} \rightarrow A(t) \in SO(3)$ in $SO(3)$. The cotangent bundle $P = T^*M$ of $M$ (Remark 2.3.2 of [Nab5] or Section 3.2 of [BG]) is called the **phase space** and the points $(x, \eta)$ in it represent **states** of the system. Here $x$ is a point in $M$ and $\eta \in T^*_x(M)$ is a **covector** at $x$, that is, an element of the dual of the tangent space $T_x(M)$ to $M$ at $x$.

**Remark A.3.1.** In the Lagrangian picture described in Section A.2 the states of the system are described by points $(x, v_x)$ in the tangent bundle $TM$ of $M$. Then $x$ represents a possible configuration of the particles and $v_x$ a possible rate of change of the configuration. Locally, in coordinates, $(x, v_x) = (q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n)$ and the transition from $TM$ to $T^*M$ amounts to replacing the $\dot{q}^i$ by the conjugate momenta $p_i = \frac{\partial L}{\partial \dot{q}^i}$ (see Section 2.3 of [Nab5]). We trust that the context (Lagrangian or Hamiltonian) will always make it clear whether the word “state” refers to a point in $TM$ or $T^*M$; these should be thought of simply as two different ways of describing the “physical state” of some underlying mechanical system.

$T^*M$ admits a canonical 1-form $\theta$ defined in the following way. For any $(x, \eta) \in T^*M$, with $x \in M$ and $\eta \in T^*_x(M)$, $\theta_{(x,\eta)}$ is the linear functional defined on $T_{(x,\eta)}(T^*M)$ by $\theta_{(x,\eta)} = \eta \circ \pi_{(x,\eta)}$, where $\pi$ is the projection of $T^*M$ onto $M$ and $\pi_{(x,\eta)}$ is its derivative at $(x, \eta)$. Near each point of $T^*M$ there are local coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$, called **canonical coordinates**, relative to which $\theta = p_1dq^1 = p_1dq^1 + \cdots + p_ndq^n$. The 2-form $\omega = -d\theta$ is closed ($d\omega = 0$) and nondegenerate $\iota_X\omega = 0 \Rightarrow X = 0$, where $\iota_X\omega$ is the contraction of $\omega$ with the smooth vector field $X$ on $T^*M$, defined by $\iota_X\omega(Y) = \omega(X, Y)$ for any smooth vector field $Y$ on $T^*M$. $\omega$ is called the **canonical symplectic form** on $T^*M$ and, in canonical coordinates, it is given by $\omega = dq^i \wedge dp_i$.

One selects a distinguished, smooth real-valued function $H : T^*M \rightarrow \mathbb{R}$, called the **Hamiltonian** and representing the total energy of the system being modeled, and defines from it and the symplectic form $\omega$ a vector field $X_H$ on $T^*M$, called the **Hamiltonian vector field**, by the requirement that $\iota_{X_H}\omega = dH$. In canonical coordinates, $X_H = (\partial H/\partial q_i)p_i - (\partial H/\partial p_i)q_i$. The state of the system is then assumed to evolve with time from some initial state along the inte-
gral curve of $X_H$ through this initial state. Consequently, the evolution of the state $(q^1(t), \ldots, q^n(t), p_1(t), \ldots, p_n(t))$ with time satisfies Hamilton’s equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \ i = 1, \ldots, n.$$ 

The Hamiltonian flow, that is, the flow of the Hamiltonian vector field $X_H$, therefore contains all of the information about the evolution of the state of the system.

**Remark A.3.2.** Depending on the nature of $H$, Hamilton’s equations may have solutions defined only on some finite interval of $t$ values so the Hamiltonian flow may be only a local flow.

**Example A.3.1.** (Particle Motion in $\mathbb{R}^n$) The standard example models a single particle of mass $m > 0$ moving in $\mathbb{R}^3$ under the influence of some conservative force $\mathbf{F} = -\nabla V$. The configuration space is $M = \mathbb{R}^3$ and we denote by $(q^1, q^2, q^3)$ its standard Cartesian coordinates. Physically, $M$ is identified with the space of possible positions of the particle. Since $\mathbb{R}^3$ is contractible, its cotangent bundle can be identified with $T^* M = \mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$. The Hamiltonian is to be the total energy of the particle, that is, the sum of its kinetic and potential energies, and we would like to describe this in canonical coordinates $(q^1, q^2, q^3, p_1, p_2, p_3)$.

**Remark A.3.3.** In Example A.2.1 we saw that the difference of the kinetic and potential energies is the Lagrangian $L$ for this system and that the momentum conjugate to $q^i$ is $\dot{q}^i = \frac{\partial L}{\partial \dot{q}^i} = m\dot{q}^i$. In light of Remark A.3.1 we have $p_i = m\dot{q}^i$ so the particle’s kinetic energy can be written $\frac{1}{2m}p_i^2$.

The particle’s potential energy due to the influence of the conservative force $-\nabla V(q)$ is just $V(q)$. The Hamiltonian $H(q, p)$ is the sum of these.

$$H(q, p) = \frac{1}{2m}\|p\|^2 + V(q) = \frac{1}{2m}(p_1^2 + p_2^2 + p_3^2) + V(q^1, q^2, q^3).$$

Relative to these canonical coordinates on $\mathbb{R}^6$ the canonical symplectic form is

$$\omega = dq^i \wedge dp_i = dq^1 \wedge dp_1 + dq^2 \wedge dp_2 + dq^3 \wedge dp_3$$

and the Hamiltonian vector field is therefore

$$X_H(q, p) = (\partial H/\partial p_i) \partial_{q^i} - (\partial H/\partial q^i) \partial_{p_i} = \frac{1}{m}p_i \partial_{q^i} - \frac{\partial V}{\partial q^i} \partial_{p_i}.$$ 

From this we read off Hamilton’s equations

$$\dot{q}^i = \frac{1}{m}p_i \quad \text{and} \quad \dot{p}_i = -\frac{\partial V}{\partial q^i}, \ i = 1, 2, 3.$$
Notice that substituting the first of these into the second gives Newton’s Second Law
\[ m\ddot{q}^i = - \frac{\partial V}{\partial q^i}, \quad i = 1, 2, 3. \]

There is an obvious generalization of all of this to $\mathbb{R}^n$ for any $n = 1, 2, 3, \ldots$ and we will record one particularly important example.

Take $n = 1$ so that the configuration space is $M = \mathbb{R}$ and the phase space is $T^*M = \mathbb{R}^2$. Writing $q^1 = q$ and $p_1 = p$ for the canonical coordinates and taking $V(q) = \frac{m\omega^2}{2}q^2$, where $\omega$ is a positive constant, we obtain the Hamiltonian
\[ H(q, p) = \frac{1}{2m}p^2 + \frac{m\omega^2}{2}q^2 \]
for the classical harmonic oscillator of mass $m$ and natural frequency $\omega$ (see Chapter 1 of [Nab5]). The Hamiltonian vector field is
\[ X_H(q, p) = \frac{1}{m}p \frac{\partial}{\partial q} - m\omega^2 q \frac{\partial}{\partial p} \]
and Hamilton’s equations are
\[ \dot{q} = \frac{1}{m}p \quad \text{and} \quad \dot{p} = -m\omega^2 q. \]
These combine to give
\[ \ddot{q} + \omega^2 q = 0 \]
which is easily solved to obtain
\[ q(t) = A \cos (\omega t + \varphi), \]
where $A$ is a non-negative constant, called the amplitude of the oscillator, and $\varphi$ is a real constant called the phase. Notice that, because the Hamiltonian is quadratic, Hamilton’s equations are linear so the integral curves are defined for all $t$ and the flow is global.

The essential data of Hamiltonian mechanics as we have discussed it thus far consists of the phase space $T^*M$ with its canonical symplectic form $\omega$ and a distinguished real-valued function $H$ on phase space (the canonical 1-form $\theta$ played only a supporting role in that it gave rise to $\omega$). The rest of the formalism, which we now review, is all derived from these (all of the details are available in Section 2.3 of [Nab5]).

Each smooth, real-valued function $f$ on the phase space $T^*M$ is called a classical observable. The space $C^\infty(T^*M; \mathbb{R})$ of smooth, real-valued functions on $T^*M$, with its usual real vector space structure and pointwise multiplication, is referred to as the algebra of classical observables. Each $f \in C^\infty(T^*M; \mathbb{R})$ has a symplectic gradient
$X_f$ defined by the requirement that $\iota_{X_f} \omega = df$.

Note: The symplectic gradient of $f$ is also called the Hamiltonian vector field of $f$, but we will reserve this term for the symplectic gradient $X_H$ of the Hamiltonian $H$.

In canonical coordinates, $X_f = (\frac{\partial f}{\partial p_i}) \partial q^i - (\frac{\partial f}{\partial q^i}) \partial p_i$. Each $X_f$ preserves the symplectic form $\omega$ in the sense that

$L_{X_f} \omega = 0$,

where $L_{X_f}$ is the Lie derivative with respect to $X_f$. This follows from Cartan’s “magic” formula for the Lie derivative since

$L_{X_f} \omega = (d\iota_{X_f} + \iota_{X_f} d) \omega = d(df) + \iota_{X_f}(d\omega) = 0 + \iota_{X_f}(0) = 0$.

Equivalently, $\varphi_t^* \omega = \omega$ for every $\varphi_t$ in the (possibly local) 1-parameter group of diffeomorphisms of $X_f$.

For $f, g \in C^\infty(T^*M; \mathbb{R})$ we define their Poisson bracket $\{f, g\}$ by $\{f, g\} = \omega(X_f, X_g)$. One finds that, in canonical coordinates,

$\{f, g\} = \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i}$

and that the Lie bracket of two symplectic gradient vector fields $X_f$ and $X_g$ is the symplectic gradient of the Poisson bracket of $f$ and $g$, that is,

$[X_f, X_g] = X_{\{f, g\}} \quad \forall f, g \in C^\infty(T^*M; \mathbb{R})$.

The Poisson bracket provides $C^\infty(T^*M; \mathbb{R})$ with the structure of a Lie algebra since

$\{\cdot, \cdot\} : C^\infty(T^*M; \mathbb{R}) \times C^\infty(T^*M; \mathbb{R}) \to C^\infty(T^*M; \mathbb{R})$

is $\mathbb{R}$-bilinear, skew-symmetric

$\{g, f\} = -\{f, g\} \quad \forall f, g \in C^\infty(T^*M; \mathbb{R})$,

and satisfies the Jacobi Identity

$\{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} = 0 \quad \forall f, g, h \in C^\infty(T^*M; \mathbb{R})$.

Moreover, $C^\infty(T^*M; \mathbb{R})$ is a Poisson algebra since the Poisson bracket satisfies the Leibniz Rule

$\{f, gh\} = \{f, g\} h + g\{f, h\} \quad \forall f, g, h \in C^\infty(T^*M; \mathbb{R})$.

In terms of the Poisson bracket Hamilton’s equations take the form
A.3 Finite-Dimensional Hamiltonian Mechanics

\[ \dot{q}^i = \{q^i, H\} \quad \text{and} \quad \dot{p}_i = \{p_i, H\}, \quad i = 1, \ldots, n. \]

Moreover, for any \( f, g \in C^\infty(T^*M; \mathbb{R}) \),
\[ \{f, g\} = X_q[f] = L_{X_q}f, \]
so, in particular,
\[ \{f, H\} = L_{X_H}f \]
is the rate of change of \( f \) along the integral curves of the Hamiltonian, that is, along the trajectories of the system. Consequently, \( f \) is conserved (constant along trajectories) if and only if it Poisson commutes with the Hamiltonian, meaning \( \{f, H\} = 0 \). In particular, the Hamiltonian itself is conserved and this is the form that conservation of energy takes in Hamiltonian mechanics.

We see then that conservation laws appear to arise rather differently in the Hamiltonian picture than in the Lagrangian and one might wonder if there is an analogue of Noether’s Theorem A.2.1 relating conserved quantities to “symmetries” in Hamiltonian mechanics. There is indeed, but the real beauty and significance of the result will not be so apparent if we continue to restrict our attention to the rather special case we have been considering to this point (the case in which the phase space is a cotangent bundle). For this reason we will now pause to describe the much more general context in which the Hamiltonian picture lives most naturally. This done, we will return to the question of symmetries and conserved quantities.

The structure we have been discussing was built from essentially just three basic ingredients: a phase space and, defined on it, a nondegenerate, closed 2-form and a distinguished smooth, real-valued function. As a result it is a simple matter to describe a very general, abstract mathematical structure that encompasses the Hamiltonian picture of classical mechanics as a special case. This is worth doing for many reasons related to quantization, representation theory, geometry, topology, and as a stepping stone to the infinite-dimensional version required to accommodate classical field theory. Our description will be relatively brief, but for those who wish to see more of this we might recommend [Bern], the article Introduction to Lie Groups and Symplectic Geometry by Robert Bryant in [FU], [AM], [AMR], [Arn2], [GS1], and [Ch].

We begin with a smooth manifold \( P \) of dimension \( 2n \) (the reason for assuming the dimension is even will be clear momentarily); for simplicity we will assume also that \( P \) is connected. \( P \) will be called the phase space. A symplectic form on \( P \) is a 2-form \( \omega \) on \( P \) that is closed (\( d\omega = 0 \)) and nondegenerate (\( \omega \wedge \omega = 0 \Rightarrow V = 0 \)). The pair \( (P, \omega) \) is called a symplectic manifold. Bilinearity and nondegeneracy imply that \( \omega \) determines an isomorphism \( \omega^\flat \) from the space \( \mathcal{X}(P) \) of smooth vector fields on \( P \) to the space \( \Omega^1(P) \) of smooth, real-valued 1-forms on \( P \) defined by
\[ \omega^\flat(X) = i_X\omega \quad \forall X \in \mathcal{X}(P). \]
The inverse of $\omega^\flat$ is denoted $\omega^\sharp : \Omega(P) \rightarrow \mathcal{X}(P)$.

**Remark A.3.4.** If a manifold $P$ has a symplectic form $\omega$ defined on it, then $P$ is necessarily even dimensional. The reason is as follows. Suppose dim $P = m$. For any $p \in P$, $\omega_p$ can be identified with a skew-symmetric $m \times m$ matrix $(\omega_p)$ and nondegeneracy implies that this matrix is nonsingular. But then $\det(\omega_p) = \det((\omega_p)^T) = (-1)^m \det (\omega_p)$ by skew-symmetry and this is possible only if $m$ is even.

$P$ must also be orientable since the nondegeneracy of $\omega$ implies that $\omega^\sharp = \omega \wedge \cdots \wedge \cdot$. $\wedge \omega$ is a nonzero 2-form on the 2$n$-dimensional manifold $P$, that is, a volume form, and the existence of a volume form is equivalent to orientability (Theorem 4.3.1 of [Nab3]). However, not every orientable, even dimensional manifold admits a symplectic form. An example is the 4-sphere $S^4$ and the reason is topological. Indeed, a symplectic form $\omega$ on $S^4$, being a closed 2-form, represents an element $[\omega]$ of the second de Rham cohomology group $H^2_{deRham}(S^4)$. This element would have to be nonzero since $[\omega] = 0 \Rightarrow [\omega]^2 = [\omega^2] = 0$ and from this one would obtain $\int_{S^4} \omega^2 = 0$ which contradicts the fact that $\omega^2$ is a volume form on $S^4$. However, $H^2_{deRham}(S^4)$ is the trivial group (Exercise 5.4.3 of [Nab3]) so such an $\omega$ cannot exist.

**Example A.3.2.** Let $M$ be any smooth, $n$-dimensional manifold. Then the cotangent bundle $T^*M$ is a smooth, 2$n$-dimensional manifold which, as we have seen, admits a (canonical) symplectic form $\omega$. Thus, $(T^*M, \omega)$ is a symplectic manifold. Notice that, for this example, $\omega$ is not only closed, but also exact ($\omega = d(\theta)$, where $\theta$ is the canonical 1-form on $T^*M$). From the perspective of de Rham cohomology this means that the 2-form $\omega$ is cohomologically trivial. This is certainly not the case for a general symplectic form. Indeed, the ideas we just used to show that $S^4$ does not admit a symplectic structure show also that the symplectic form on any compact symplectic manifold is cohomologically nontrivial, that is, represents a nonzero second cohomology class. There are, incidentally, lots of compact symplectic manifolds (see Example A.3.4).

**Example A.3.3.** Let $\mathcal{V}$ be a 2$n$-dimensional real vector space (like $\mathbb{R}^{2n}$ or the tangent space to a 2$n$-dimensional manifold). Any basis $\{e_1, \ldots, e_{2n}\}$ for $\mathcal{V}$ determines a natural topology and differentiable structure for $\mathcal{V}$. Specifically, if $\{e^1, \ldots, e^{2n}\}$ is the dual basis, then $p \in \mathcal{V} \mapsto (x^1, \ldots, x^{2n}) = (e^1(p), \ldots, e^{2n}(p)) \in \mathbb{R}^{2n}$ is a bijection. We supply $\mathcal{V}$ with the unique topology for which this map is a homeomorphism. A different choice of basis determines the same topology so these homeomorphisms are charts on the topological space $\mathcal{V}$. These charts overlap smoothly and so determine a differentiable structure on $\mathcal{V}$. The tangent space $T_p(\mathcal{V})$ at any $p \in \mathcal{V}$ is naturally identified with $\mathcal{V}$ itself. Specifically, each $v_p \in T_p(\mathcal{V})$ is the velocity vector to the curve $t \mapsto p + tv$ for some $v \in \mathcal{V}$ and $v_p \mapsto v$ is an isomorphism. From now on we will make this identification without further comment.

Now, let $S : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ be a nondegenerate, skew-symmetric, bilinear form on $\mathcal{V}$. For example, if $\langle \cdot, \cdot \rangle$ denotes the usual positive definite inner product on $\mathbb{R}^n$ and if we regard $\mathbb{R}^{2n}$ as $\mathbb{R}^n \times \mathbb{R}^n$, then
A.3 Finite-Dimensional Hamiltonian Mechanics

\[ S((x_1, y_1), (x_2, y_2)) = \langle x_1, y_2 \rangle - \langle y_1, x_2 \rangle \]

defines such a bilinear form. Any such \( S \) determines a (constant) symplectic form \( \omega \) on the manifold \( V \) by defining, at each \( p \in V \),

\[ \omega_p(v_p, w_p) = S(v, w). \]

It would seem then that one can produce a wide variety of symplectic forms on \( V \) by simply making various choices for \( S \). These hopes are dashed, however, by a theorem in linear algebra according to which these are, up to a change of linear coordinates, all the same. The following is Proposition 1.3(ii) of [Bern].

**Theorem A.3.1.** (Linear Darboux Theorem) Let \( V \) be a \( 2n \)-dimensional real vector space, \( S : V \times V \to \mathbb{R} \) a nondegenerate, skew-symmetric, bilinear form on \( V \), and \( \omega \) the corresponding symplectic form on \( V \). Then there exists a basis \( \{e_1, \ldots, e_n, e_{n+1}, \ldots, e_{2n}\} \) for \( V \) with corresponding linear coordinates \( (q^1, \ldots, q^n, p_1, \ldots, p_n) \) such that

\[ \omega = dq^i \wedge dp_i = -d(p_i dq^i). \]

It is remarkable that this result extends locally to arbitrary symplectic manifolds. For a very detailed proof of the following theorem of Darboux see Section 2.2 of [Bern].

**Theorem A.3.2.** (Darboux Theorem) Let \( (P, \omega) \) be a symplectic manifold of dimension \( 2n \). Then, at each point in \( P \), there exists a local coordinate neighborhood \( U \) with coordinates \( (q^1, \ldots, q^n, p_1, \ldots, p_n) \) such that

\[ \omega = dq^i \wedge dp_i = -d(p_i dq^i). \]

**Remark A.3.5.** The essential content of the Darboux Theorem is that all symplectic manifolds of the same dimension are locally the same. This contrasts rather markedly with Riemannian manifolds of the same dimension, for which the local structure is determined by curvature. This gives symplectic geometry and topology quite a different flavor from classical differential geometry and topology (see, for example, [MS]).

**Example A.3.4.** Let \( P \) be any orientable smooth surface (that is, orientable 2-dimensional manifold) and \( \omega \) a volume (area) form on \( P \). Then \( \omega \) is a 2-form and it is closed because every 2-form on a 2-manifold is closed. It is also nondegenerate. To see this we assume that \( V \) is a vector field on \( P \) for which \( \iota_V \omega = 0 \) and will show that \( V(p) = 0 \) at each \( p \in P \). Since \( \omega \) determines an orientation for \( P \) there exists, on a neighborhood \( U \) of \( p \), a chart with coordinates \( (x, y) \) for which \( \omega(\partial_x, \partial_y) > 0 \) (see Theorem 4.3.1 of [Nab3]). Then \( \omega(\partial_y, \partial_x) < 0 \) and \( \omega(\partial_x, \partial_x) = \omega(\partial_y, \partial_y) = 0 \) on \( U \).
Now write $V = V_x \partial_x + V_y \partial_y$ for some smooth functions $V_x$ and $V_y$ on $U$. Then, by assumption,

$$0 = \omega(V_x \partial_x + V_y \partial_y, \partial_x) = V_x \omega(\partial_x, \partial_x)$$

at every point of $U$. Consequently, $V_x = 0$ on $U$. Similarly, $V_y = 0$ on $U$. In particular, $V(p) = V_x(p)\partial_x(p) + V_y(p)\partial_y(p) = 0$ as required. Consequently, $\omega$ is a symplectic form on $P$ and so every orientable surface is a symplectic 2-manifold.

Since lots of these are compact ($S^2$ for example) we have fulfilled our promise to exhibit examples of compact symplectic manifolds (Example A.3.2).

Now suppose $(P, \omega)$ is an arbitrary symplectic manifold of dimension $2n$. The Darboux Theorem implies that, at each point of $P$, there exists a local coordinate neighborhood $U$ with coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ such that, on $U$,

$$\omega = dq^i \wedge dp_i.$$

We call these canonical coordinates. The algebra $C^\infty(P; \mathbb{R})$ of smooth real-valued functions on $P$ with its usual real vector space structure and pointwise multiplication is called the algebra of classical observables and each element of it is called a classical observable. Select some element $H$ of $C^\infty(P; \mathbb{R})$, christen it the Hamiltonian and think of it as the total energy of a physical system whose phase space is being modeled by $(P, \omega)$. The pair $((P, \omega), H)$ is then called a Hamiltonian system.

Since $\omega$ is nondegenerate any covector in $T^*_p(P)$ is $\omega_p(v_p, \cdot)$ for some tangent vector $v_p \in T_p(P)$ and every 1-form on $P$ is $\omega(V, \cdot) = \iota_V \omega$ for some smooth vector field $V$ on $P$. In particular, any smooth real-valued function $f$ on $P$ has a symplectic gradient $X_f$ defined by the requirement that $\iota_{X_f} \omega = df$, that is, $\omega(X_f, \cdot) = df(\cdot)$. The symplectic gradient $X_H$ of the Hamiltonian is called the Hamiltonian vector field.

Remark A.3.6. Again we point out that $X_f$ is often called the Hamiltonian vector field of $f$.

Exercise A.3.1. Show that, in local canonical coordinates,

$$X_f = (\partial f / \partial p_i) \partial_{q^i} - (\partial f / \partial q^i) \partial_{p_i}.$$

In particular, the integral curves of the Hamiltonian vector field $X_H$ are locally given by functions $(q^1(t), \ldots, q^n(t), p_1(t), \ldots, p_n(t))$ that satisfy Hamilton’s equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \quad i = 1, \ldots, n$$

and the rate of change of any classical observable $f \in C^\infty(P; \mathbb{R})$ along these integral curves is locally given by
A.3 Finite-Dimensional Hamiltonian Mechanics

\[
X_H[f] = L_{X_H} = \frac{\partial H}{\partial p_i} \frac{\partial f}{\partial q^i} - \frac{\partial H}{\partial q_i} \frac{\partial f}{\partial p^i}.
\]

More generally, we define, for any \( f, g \in C^\infty(P; \mathbb{R}) \), the Poisson bracket of \( f \) and \( g \) determined by \( \omega \) by

\[
\{ f, g \} = \omega(X_f, X_g)
\]

and find that, locally in canonical coordinates,

\[
\{ f, g \} = \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i}.
\]

Exercise A.3.2. Let \((P, \omega)\) be an arbitrary symplectic manifold and let \((q, p) = (q^1, \ldots, q^n, p_1, \ldots, p_n)\) be canonical coordinates on some open neighborhood \( U \) in \( P \). Let \( \omega_U \) denote the restriction of \( \omega \) to \( U \), that is, the pullback of \( \omega \) to \( U \) by the inclusion map \( U \hookrightarrow P \). Show that \((U, \omega_U)\) is a symplectic manifold and that the canonical coordinate functions \( q^1, \ldots, q^n, p_1, \ldots, p_n \in C^\infty(U; \mathbb{R}) \) satisfy the classical canonical commutation relations

\[
\{ q^i, q^j \} = \{ p_i, p_j \} = 0 \quad \text{and} \quad \{ q^i, p_j \} = \delta_{ij}, \quad i, j = 1, \ldots, n, \quad (A.8)
\]

where \( \delta_{ij} \) is the Kronecker delta.

One can now show that, in this more general context, all of the fundamental properties that we enumerated for the cotangent bundle remain valid.

Remark A.3.7. The proofs given in Section 2.3 of [Nab5] for the cotangent bundle are intentionally phrased in such a way as to carry over verbatim to arbitrary symplectic manifolds.

Specifically, one can prove all of the following.

\[ L_{X_f} \omega = 0 \quad \forall f \in C^\infty(P; \mathbb{R}) \]

\[ [X_f, X_g] = X_{\{f, g\}} \quad \forall f, g \in C^\infty(P; \mathbb{R}) \]

Furthermore

\[ \{ , \} : C^\infty(P; \mathbb{R}) \times C^\infty(P; \mathbb{R}) \rightarrow C^\infty(P; \mathbb{R}) \]

is \( \mathbb{R} \)-bilinear, skew-symmetric

\[ \{ g, f \} = -\{ f, g \} \quad \forall f, g \in C^\infty(P; \mathbb{R}), \]

satisfies the Jacobi Identity
\[ \{ f, \{ g, h \} \} + \{ h, \{ f, g \} \} + \{ g, \{ h, f \} \} = 0 \quad \forall f, g, h \in C^\infty(P; \mathbb{R}), \]

and the Leibniz Rule

\[ \{ f, gh \} = \{ f, g \} h + g \{ f, h \} \quad \forall f, g, h \in C^\infty(P; \mathbb{R}). \]

Moreover, in terms of the Poisson bracket, Hamilton’s equations take the form

\[ \dot{q}^i = \{ q^i, H \} \quad \text{and} \quad \dot{p}_i = \{ p_i, H \}, \quad i = 1, \ldots, n \]

and \( f \in C^\infty(P; \mathbb{R}) \) is conserved (constant along the integral curves of the Hamiltonian vector field) if and only if it Poisson commutes with the Hamiltonian, that is, if and only if

\[ \{ f, H \} = 0. \]

In particular, the Hamiltonian itself (that is, the total energy) is clearly conserved since \( \{ H, H \} = 0 \) follows from the skew-symmetry of the Poisson bracket. Moreover, it follows from the Jacobi identity that the Poisson bracket of two conserved quantities is also conserved. Indeed,

\[ \{ f, H \} = \{ g, h \} = 0 \Rightarrow \{ \{ f, g \}, H \} = \{ g, \{ H, f \} \} + \{ f, \{ g, H \} \} = \{ g, 0 \} + \{ f, 0 \} = 0. \]

More generally, even if \( f \) is not conserved, the Poisson bracket keeps track of how it evolves with the system in the sense that, along an integral curve of \( X_H \),

\[ \frac{df}{dt} = \{ f, H \} \quad \text{(A.9)} \]

(because \( X_H[f] = -\{ H, f \} = \{ f, H \} \)).

Now we are prepared to return to the question of symmetries and conservation laws (see page 137). We begin with a few general definitions. Let \( (P_1, \omega_1) \) and \( (P_2, \omega_2) \) be two symplectic manifolds. A map \( F : P_1 \to P_2 \) is called a symplectic diffeomorphism, or symplectomorphism, or, in the physics literature, a canonical map if it is a diffeomorphism of \( P_1 \) onto \( P_2 \) that carries \( \omega_1 \) onto \( \omega_2 \) in the sense that

\[ F^* \omega_2 = \omega_1. \]

A smooth vector field \( X \) on a symplectic manifold \( (P, \omega) \) is called a symplectic vector field if it preserves the symplectic form in the sense that

\[ L_X \omega = 0. \]

If \( X \) is complete this is equivalent to the requirement that each \( \varphi_t, t \in \mathbb{R}, \) in the 1-parameter group of diffeomorphisms determined by \( X \) is a symplectic diffeomorphism. A vector field \( X \) on \( (P, \omega) \) is said to be Hamiltonian if it is the symplectic gradient of some \( f \in C^\infty(P; \mathbb{R}) \). Every Hamiltonian vector field is therefore also a
symplectic vector field. That the converse is generally not true follows from the next exercise.

**Exercise A.3.3.** Let $X$ be a smooth vector field on $(P, \omega)$. Prove each of the following.

1. $X$ is symplectic if and only if the 1-form $\iota_X \omega$ is closed.
2. $X$ is Hamiltonian if and only if the 1-form $\iota_X \omega$ is exact.

When the first de Rham cohomology group of $P$ is trivial (for example, when $P = \mathbb{R}^{2n}$), every closed 1-form is exact so a vector field $X$ on $(P, \omega)$ is symplectic if and only if it is Hamiltonian. More generally, since $P$ is locally diffeomorphic to $\mathbb{R}^{2n}$ any symplectic vector field on $P$ is locally Hamiltonian.

A symmetry of the Hamiltonian system $((P, \omega), H)$ is a symplectic diffeomorphism $F : P \to P$ of $P$ onto itself that preserves the Hamiltonian $H$ in the sense that

$$F^* H = H,$$

that is,

$$H \circ F = H.$$

Thus, a symmetry is a diffeomorphism of phase space that preserves both the symplectic form and the Hamiltonian. For the infinitesimal version we proceed as in the Lagrangian case (see page 126). A smooth, complete vector field $X$ on $P$ is said to be an infinitesimal symmetry of the Hamiltonian system $((P, \omega), H)$ if each $\varphi_t, t \in \mathbb{R}$, in the 1-parameter group of diffeomorphisms of $X$ is a symmetry of $((P, \omega), H)$; if $X$ is not complete then one modifies the definition exactly as in the Lagrangian case (page 126).

Infinitesimal symmetries generally arise from group actions of the following type. Let $G$ be a (matrix) Lie group with Lie algebra $\mathfrak{g}$ and suppose $\sigma : G \times P \to P$, $\sigma(g, x) = \sigma_{\xi}(x) = g \cdot x$, is a smooth left action of $G$ on $P$. If each of the diffeomorphisms $\sigma_g : P \to P, g \in G$, is a symmetry of the Hamiltonian system $((P, \omega), H)$, then we refer to $G$ as a symmetry group of $((P, \omega), H)$. In this case each nonzero element $\xi$ of $\mathfrak{g}$ gives rise to an infinitesimal symmetry $X_\xi$ defined at each $x \in P$ by

$$X_\xi(x) = \frac{d}{dt}(e^{t \xi} \cdot x) \bigg|_{t=0}.$$  

In order to write down a Hamiltonian version of Noether’s Theorem A.2.1 we introduce an idea that is fundamental to modern symplectic geometry. For this we denote by $\mathfrak{g}^*$ the vector space dual of the Lie algebra $\mathfrak{g}$. This is a finite-dimensional real vector space so it has a natural topology and differentiable structure. Consider a smooth map
\[ \mu : P \to g^*. \]

Then for every \( x \in P \), \( \mu(x) \) is a real-valued linear map on \( g \)

\[ \mu(x) : g \to \mathbb{R} \]

so that \( \langle \mu(x)(\xi) \rangle \) is a real number for every \( \xi \in g \). We want to fix \( \mu \) and \( \xi \) and regard this as a real-valued function on \( P \), that is, we define

\[ \langle \mu, \xi \rangle : P \to \mathbb{R} \]

by

\[ \langle \mu, \xi \rangle(x) = \mu(x)(\xi). \]

**Note:** The notation is meant to suggest the natural pairing of \( g \) and \( g^* \) so that one should think of \( \langle \mu, \xi \rangle \) as being defined by

\[ \langle \mu, \xi \rangle(x) = \langle \mu(x), \xi \rangle. \]

The map \( \langle \mu, \xi \rangle \) is smooth so it has a symplectic gradient \( X_{\langle \mu, \xi \rangle} \). We will say that \( \mu \) is a moment map, also called a momentum map, for the Hamiltonian system \( ((P, \omega), H) \) with respect to the symmetry group \( G \) if

\[ X_{\langle \mu, \xi \rangle} = X_{\xi} \]

for every \( \xi \in g \). The Hamiltonian version of Noether’s Theorem A.2.1 asserts that if \( \mu \) is a moment map for the Hamiltonian system \( ((P, \omega), H) \) with respect to \( G \), then \( \langle \mu, \xi \rangle \) is conserved for every \( \xi \in g \).

**Theorem A.3.3.** (Noether’s Theorem: Hamiltonian Version) Let \( ((P, \omega), H) \) be a Hamiltonian system, \( G \) a symmetry group of \( ((P, \omega), H) \) with Lie algebra \( g \), and \( \mu : P \to g^* \) a moment map for \( ((P, \omega), H) \) with respect to \( G \). Then, for every \( \xi \in g \), the function

\[ \langle \mu, \xi \rangle : P \to \mathbb{R} \]

defined by

\[ \langle \mu, \xi \rangle(x) = \mu(x)(\xi) \]

for every \( x \in P \) is constant along the integral curves of the Hamiltonian vector field \( X_H \). In particular, each \( \langle \mu, \xi \rangle \) Poisson commutes with the Hamiltonian \( H \).

\[ [\langle \mu, \xi \rangle, H] = 0 \]
A.3 Finite-Dimensional Hamiltonian Mechanics

Proof. Let $\gamma(t)$ be an integral curve of $X_H$. Then, by definition, $\dot{\gamma}(t) = X_H(\gamma(t))$ for each $t$. We compute, for each fixed $\xi \in \mathfrak{g}$,

$$
\frac{d}{dt}[(\mu, \xi)(\gamma(t))] = d(\mu, \xi)_{\gamma(t)}(\dot{\gamma}(t))
$$

$$
= d(\mu, \xi)_{\gamma(t)}(X_H(\gamma(t)))
$$

$$
= (\iota_{X_H} \omega)_{\gamma(t)}(X_H(\gamma(t)))
$$

$$
= \omega(X_{\mu, \xi}, \cdot)_{\gamma(t)}(X_H(\gamma(t)))
$$

$$
= \omega(X_{\mu, \xi}, \cdot)_{\gamma(t)}(X_H(\gamma(t)))
$$

$$
= -\omega_{\gamma(t)}(X_H(\gamma(t)), X_H(\gamma(t)))
$$

$$
= -dH_{\gamma(t)}(X_{\xi}(\gamma(t)))
$$

$$
= -dH_{\gamma(t)}\left(\frac{d}{ds}(e^{s \xi} \cdot \gamma(t))\big|_{s=0}\right)
$$

$$
= -\frac{d}{ds}(H(e^{s \xi} \cdot \gamma(t)))\big|_{s=0}
$$

$$
= -\frac{d}{ds}(H(\gamma(t)))\big|_{s=0}
$$

$$
= 0
$$

as required. \qed

This result does not address the issue of actually finding moment maps or the question of their existence. As it happens a symmetry group for a Hamiltonian system need not have a moment map, although there are broad classes of such problems for which the existence of moment maps can be proved; for an introduction to this see, for example, Chapter 4 of [AM]. We will conclude this section with just one simple example that should at least clarify the origin of the terminology.

Example A.3.5. We construct a Hamiltonian system $((P, \omega), H)$ as follows. Let $P = T^*\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3$ be the tangent bundle of $\mathbb{R}^3$ with its canonical symplectic form $\omega = dq^1 \wedge dp_1 = dq^1 \wedge dp_1 + dq^2 \wedge dp_2 + dq^3 \wedge dp_3$.

For the Hamiltonian we can take any smooth map $H : T^*\mathbb{R}^3 \to \mathbb{R}$ that satisfies $H(q + a, p) = H(q, p)$ for all $(q, p) \in T^*\mathbb{R}^3$ and any $a \in \mathbb{R}^3$. One possibility is the kinetic energy Hamiltonian

$$
H(q, p) = \frac{||p||^2}{2m},
$$

where $m$ is a positive constant. Next we need to find a symmetry group for this Hamiltonian system. Take $G$ to be the additive group $\mathbb{R}^3$. We think of $G$ as the translation group of $\mathbb{R}^3$, that is, we define a left action $\sigma : G \times P \to P$ of $G$ on $P$ by
\[ \sigma(a, (q, p)) = (a + q, p) \]

for all \((a, (q, p)) \in G \times P\). Then each map \(\sigma_a : P \rightarrow P\) defined by \(\sigma_a(q, p) = (a + q, p)\) is a diffeomorphism and, by assumption,

\[ \sigma_a^* H = H. \]

Moreover,

\[ \sigma_a^* \omega = \sigma_a^*(dq^i \wedge dp_i) = \sigma_a^*(dq^i) \wedge \sigma_a^*(dp_i) = dq^i \wedge dp_i = \omega. \]

Consequently, each \(\sigma_a\) is a symmetry of the Hamiltonian system so \(G = \mathbb{R}^3\) acts as a symmetry group.

We will identify the Lie algebra of \(\mathbb{R}^3\) with \(\mathbb{R}^3\) (Exercise 1.1.1). The exponential map on the Lie algebra satisfies \(e^{t\xi} = t\xi\) for every \(\xi \in \mathfrak{g}\) and every \(t \in \mathbb{R}\) (Exercise 1.1.3). Consequently, the vector field \(X_{\xi}(q, p) = \frac{d}{dt} e^{t\xi} \cdot (q, p) \big|_{t=0}\) is given by

\[ X_{\xi} = \xi^i \frac{\partial}{\partial q^i}. \]

To find a moment map for this \(G\)-action on \(((P, \sigma), H)\) we need a smooth map \(\mu : P \rightarrow \mathfrak{g}^*\) for which

\[ X_{\mu(q)} = X_{\xi}. \]

But \(X_{\mu(q)}\) is characterized by \(\iota_{X_{\mu(q)}} \omega = d\mu(\xi)\) so what we need is

\[ \iota_{X_{\xi}} \omega = d\mu(\xi). \]

**Exercise A.3.4.** Show that \(\iota_{X_{\xi}} \omega = \iota_{X_{\xi}}(dq^i \wedge dp_i)\) is the 1-form

\[ \iota_{X_{\xi}} \omega = \xi^i dp_i = \xi^1 dp_1 + \xi^2 dp_2 + \xi^3 dp_3. \]

According to this exercise \(\iota_{X_{\xi}} \omega\) will be equal to \(d\mu(\xi)\) if we define \(\mu\) in such a way that \(\langle \mu, \xi \rangle(q, p) = \xi^1 p_1 + \xi^2 p_2 + \xi^3 p_3\) for each \(\xi\), that is,

\[ \mu(q, p) = d\mu(\xi) = \xi^i p_i. \]
A.4 Postulates of Quantum Mechanics

\[ \mu(q, p) = p \]

so that, in this case, the moment map, or momentum map, really is momentum and we conclude, as we did in Section A.2, that spatial translation symmetry implies the conservation of momentum.

What we have tried to do in this section is merely suggest that the mathematical structure of classical Hamiltonian mechanics is but a special case of the much more general and very elegant subject of symplectic geometry. We have made no attempt to capitalize on this by exploiting ideas from symplectic geometry to shed light on mechanics; this is an enormous subject and there are many superb introductions to it available (see, for example, [AM], [AMR], [Arn2] and [GS1]).

A.4 Postulates of Quantum Mechanics

Physical theories are expressed in the language of mathematics, but these mathematical models are not the same as the physical theories and they are not unique. Symplectic geometry provides one possible context in which to understand classical particle mechanics, but not the only one. The Lagrangian formulation has a rather different flavor than the Hamiltonian and the two are not equivalent, but each encompasses all of classical Newtonian mechanics and each provides its own particular insights. Appropriate mathematical contexts in which to formulate the principles of quantum mechanics can likewise be constructed in a number of different ways, but we will focus our attention on just one of these which goes back to von Neumann [v.Neu]; another approach due to Feynman is discussed in Chapter 8 of [Nab5]. Both are quite unlike anything one might naively anticipate from classical mechanics, although there are precursors in classical statistical mechanics (Sections 2.4 and 3.3 of [Nab5]). The roots of the mathematical formalism lie deep in the analysis of the physical phenomena that the theory purports to describe (Chapter 4 of [Nab5]). Section 6.2 of [Nab5] introduces nine Postulates that seem to capture at least the formal structure of quantum mechanics. Since our objective here is substantially more modest we will limit ourselves to only four of these and will discuss them in somewhat less detail. You will notice that the term *quantum system* is used repeatedly, but never defined; the same is true of the term *measurement*. It is not possible, nor would it be profitable, to try to define these precisely; they are defined by the assumptions we make about them in the postulates. Each of these postulates deserves a commentary on where it came from, what it is intended to mean, how it should be interpreted, and what objections might be raised to it. However, since an attempt was made to address these issues in Section 6.2 of [Nab5] we will content ourselves here with just a few comments on each following the precise statement of the postulate. We will conclude this section with a brief description of the examples to which we will need to refer in the main body of the text.
Postulate QM1

*To every quantum system is associated a separable, complex Hilbert space \( \mathcal{H} \). The states of the system are represented by vectors \( \psi \in \mathcal{H} \) with \( ||\psi|| = 1 \) and, for any \( c \in \mathbb{C} \) with \( |c| = 1 \), \( \psi \) and \( c\psi \) represent the same state.*

In 1900 Max Planck [Planck] appended to time-honored concepts in classical physics what we would today call an *ad hoc* “quantization condition” in order to solve the classically perplexing problem of the equilibrium distribution of electromagnetic energy in a black box (this is described in some detail in Section 4.3 of [Nab5]). For the next quarter of a century physicists devised ever more ingenious such appendages to classical physics in order to solve the equally perplexing problem of atomic structure. These efforts had some limited success, but were not directed by any underlying theoretical or mathematical model of the quantum world. This changed in 1925-26 when Heisenberg [Heis1] and Schrödinger [Schrö1] each proposed such models. On the surface these looked quite different. Heisenberg’s formalism eventually came to be known as *matrix mechanics* since it was expressed in terms of infinite arrays of complex numbers (see Section 7.1 of [Nab5]). Schrödinger’s approach, called *wave mechanics*, represented the state of a quantum system by a complex-valued wave function that was assumed to satisfy a certain partial differential equation, known today as the *Schrödinger equation*. Eventually it came to be understood that these two are both physically and mathematically equivalent (see [Cas]). Naively, this equivalence can be understood in the following way. Schrödinger’s wave function \( \psi \) is most naturally regarded as an element of \( L^2(M) \), where \( M \) is the configuration space of the classical mechanical system whose quantum counterpart is under consideration (for example, the classical 2-body problem if one is interested in the hydrogen atom). But an orthonormal basis for \( L^2 \) gives rise to an isometric isomorphism onto \( l^2 \) so anything one might want to say about complex-valued square integrable functions can equally well be said in terms of finite arrays of complex numbers. Indeed, all separable, infinite-dimensional, complex Hilbert spaces are isometrically isomorphic so, when von Neumann set himself the task of constructing a mathematically rigorous, abstract setting for quantum theory, it was more natural to formulate it in terms of an arbitrary such Hilbert space \( \mathcal{H} \). For any particular quantum system the choice of \( \mathcal{H} \) then became a matter of convenience and clarity.

The reason for assuming that the states are represented by *unit* vectors in \( \mathcal{H} \) is more subtle and will be addressed more thoroughly in Postulate QM3. Briefly, the situation is as follows. In its early years (1925-1927) quantum physics was in a rather odd position. Heisenberg formulated his matrix mechanics without knowing what a matrix is and without having a precise idea of what the entries in his rectangular arrays of complex numbers should mean physically (see Section 7.1 of [Nab5] for more on this). Nevertheless, the rules of the game as he laid them down predicted precisely the spectrum of the hydrogen atom. Schrödinger formulated a differential equation for his wave function that yielded the same predictions, but no one had any real idea what the wave function was supposed to represent (Schrödinger himself initially viewed it as a sort of “charge distribution”). It was left to Max Born, and
then Niels Bohr and his school in Copenhagen, to supply the missing conceptual basis for quantum mechanics.

Remark A.4.1. It is our good fortune that Born himself, in his Nobel Prize Lecture in 1954, has provided us with a brief and very lucid account of the evolution of his idea and we will simply refer those interested in pursuing this to [Born1]. Interestingly, Born attributes to Einstein the inspiration for the idea, although Einstein never acquiesced to its implications.

As we will see, Born postulated that Schrödinger’s wave function $\psi$ should be interpreted as a probability amplitude. In particular, for a single particle moving along a line, $|\psi(q)|^2$ is the probability density function for the particle’s position, that is, for any Borel set $S$ in $\mathbb{R}$, $\int_S |\psi(q)|^2 dq$ is the probability that a measurement of the particle’s position will result in a value in $S$. Since the particle is bound to be found somewhere in $\mathbb{R}$, $\int_{\mathbb{R}} |\psi(q)|^2 dq = 1$ so $\psi$ is a unit vector in $L^2(\mathbb{R})$. All of this will be spelled out in more detail quite soon.

Finally, we point out that, because unit vectors in $\mathcal{H}$ that differ only by a phase factor describe the same state, one can identify the state space of a quantum system with the projectivization $\mathbb{P}(\mathcal{H})$ of $\mathcal{H}$, that is, the quotient of the unit sphere in $\mathcal{H}$ by the equivalence relation that identifies two points if they differ by a complex factor of modulus one. The mathematical structure of $\mathbb{P}(\mathcal{H})$ is spelled out in more detail in Section 1.3, but for the moment we need only observe that, with the quotient topology, $\mathbb{P}(\mathcal{H})$ is a Hausdorff topological space. We will write $\mathcal{P}$ for the equivalence class containing $\psi$ and refer to it as a unit ray in $\mathcal{H}$. It is sometimes also convenient to identify the state represented by the unit vector $\psi$ with the operator $P_\psi$ that projects $\mathcal{H}$ onto the 1-dimensional subspace of $\mathcal{H}$ spanned by $\psi$ (which clearly depends only on the state and not on the unit vector representing it). In all candor, however, it is customary to be somewhat loose with the terminology and speak of “the state $\psi$” when one really means “the state $\mathcal{P}$”, or “the state $P_\psi$”. Since this is almost always harmless, we will generally adhere to the custom.

Postulate QM2

*For a quantum system with Hilbert space $\mathcal{H}$, every observable is identified with a (generally unbounded) self-adjoint operator $A : \mathcal{D}(A) \to \mathcal{H}$ on $\mathcal{H}$ and any possible outcome of a measurement of the observable is a real number that lies in the spectrum $\sigma(A)$ of $A$.*

Classically, one thinks of an observable associated with a physical system as something specified by a real number that one can measure. For a single particle moving in space one might think of a coordinate of the particle’s position, a component of its momentum, or its total energy. It would be more accurate, however, to fully identify an observable with a specific measurement procedure not only because such things as position, momentum, and energy are defined operationally in
physics by specifying how they are to be measured, but also because familiar classical concepts such as these do not always transition well into the quantum realm. The problem of measurement in quantum mechanics is very subtle and, as we will see in Postulate QM3, quantum theory makes no predictions whatsoever regarding the outcome of any single measurement even when the state of the system being measured is known with certainty. Repeated measurements on identical systems in the same state need not give the same result. The results of these measurements are constrained, however, by Postulate QM2 because every quantum observable has a specific set of possible measured values, that is, $\sigma(A)$. It turns out, for example, that a quantum harmonic oscillator must have a total energy that lies in a countable, discrete set of real numbers tending to infinity (see Section 7.4 of [Nab5]).

In the model we are in the process of constructing we require a functional analytic object associated with the quantum Hilbert space $\mathcal{H}$ that encodes the essential physical content of this notion of a quantum observable. This essential content is the observable’s set of possible measured values. Now recall that a closed, symmetric operator $A$ on $\mathcal{H}$ has a spectrum $\sigma(A)$ that consists entirely of real numbers. This suggests that one might try to identify a quantum observable such as the total energy with some appropriately chosen closed, symmetric operator whose spectrum is equal to (or, at least, contains) the set of possible measured values. Such an idea has analogues in classical physics. In continuum mechanics, for example, the stress tensor of a 3-dimensional material body is a symmetric linear operator on $\mathbb{R}^3$ whose matrix in any Cartesian coordinate system is obtained from the stress components within the body in directions perpendicular to the coordinate planes and whose eigenvalues are the principal stresses, that is, those that are independent of the coordinate system (see [Gurtin]).

Postulate QM2, however, specifies that an observable is represented by a self-adjoint operator whose spectrum contains the set of possible measured values. Now, every self-adjoint operator is closed and symmetric, but the converse is not true so this is a strictly stronger requirement. The rationale behind this is largely mathematical rather than physical. The formalism of quantum mechanics depends crucially on two of the pillars of functional analysis, namely, the Spectral Theorem and Stone’s Theorem, and both of these require self-adjointness (Section 5.5 of [Nab5] or Chapter IX, Section 9, and Chapter XI, Section 6, of [Yos]). Notice that if the set of possible measured values is unbounded (as it is for the harmonic oscillator, for example), then the spectrum must be an unbounded subset of $\mathbb{R}$ and therefore the operator itself must be unbounded.

**Postulate QM3**

Let $\mathcal{H}$ be the Hilbert space of a quantum system, $\psi \in \mathcal{H}$ a unit vector representing a state of the system and $A : \mathcal{D}(A) \to \mathcal{H}$ a self-adjoint operator on $\mathcal{H}$ representing an observable. Let $E^A$ be the unique projection-valued measure on $\mathbb{R}$ associated with $A$ by the Spectral Theorem and let $\{E^A_\lambda\}_{\lambda \in \mathbb{R}}$ be the corresponding resolution of the identity. Denote by $\mu_{\psi,A}$ the probability measure on $\mathbb{R}$ that assigns to every Borel
If the state $\psi$ is in the domain of $A$, then the expected value of $A$ in state $\psi$ is

$$\langle A \rangle_\psi = \int_{\mathbb{R}} \lambda d\langle \psi, E^A_\lambda \psi \rangle = \langle \psi, A\psi \rangle$$

and its dispersion (variance) is

$$\sigma^2_\psi(A) = \int_{\mathbb{R}} (\lambda - \langle A \rangle_\psi)^2 d\langle \psi, E^A_\lambda \psi \rangle = \| (A - \langle A \rangle_\psi) \psi \|^2 = \| A\psi \|^2 - \langle A \rangle^2_\psi.$$
\( \mu_{\psi, A}, \langle A \rangle_\psi, \text{ and } \sigma^2_\psi (A) \) for various operators and states in Examples 6.2.2 - 6.2.7 of [Nab5].

Before moving on we would like to record a general result on dispersions of self-adjoint operators that is related to the uncertainty relations of quantum mechanics. The following is Lemma 6.3.1 of [Nab5].

**Lemma A.4.1.** Let \( \mathcal{H} \) be a separable, complex Hilbert space, \( A : \mathcal{D}(A) \to \mathcal{H} \) and \( B : \mathcal{D}(B) \to \mathcal{H} \) self-adjoint operators on \( \mathcal{H} \), and \( \alpha \) and \( \beta \) real numbers. Then, for every \( \psi \in \mathcal{D}([A, B]) = \mathcal{D}(AB) \cap \mathcal{D}(BA) \),

\[
\| (A - \alpha)\psi \|^2 \| (B - \beta)\psi \| ^2 \geq \frac{1}{4} \left| \langle \psi, [A, B]\psi \rangle \right|^2.
\]

Now, if \( A \) and \( B \) represent observables and \( \psi \in \mathcal{D}([A, B]) \) is a unit vector representing a state and if we take \( \alpha \) and \( \beta \) to be the corresponding expected values of \( A \) and \( B \) in this state, then it is shown in Section 6.3 of [Nab5] that Lemma A.4.1 implies

\[
\sigma^2_\psi (A) \sigma^2_\psi (B) \geq \frac{1}{4} \left| \langle \psi, [A, B]\psi \rangle \right|^2.
\]

This is called the Robertson Uncertainty Relation. It is more commonly expressed in terms of positive square roots \( \sigma_\psi (A) \) and \( \sigma_\psi (B) \) of the dispersions, that is, in terms of the standard deviations.

\[
\sigma_\psi (A) \sigma_\psi (B) \geq \frac{1}{2} \left| \langle \psi, [A, B]\psi \rangle \right|.
\]

In Section 6.3 of [Nab5] this is applied to the operators representing the position \( Q \) and momentum \( P \) of a single particle moving along a line to obtain

\[
\sigma_\psi (Q) \sigma_\psi (P) \geq \frac{\hbar}{2}, \quad (A.10)
\]

where \( \hbar = \frac{\hbar}{2\pi} \) is the reduced Planck constant. This is obviously a statistical statement about position and momentum measurements for such a particle. It is quite common, and totally incorrect, to see it identified with the famous Heisenberg Uncertainty Principle which, as Heisenberg phrased it, states that

\[
\Delta q \Delta p \geq \frac{\hbar}{2}, \quad (A.11)
\]

where \( \Delta q \) and \( \Delta p \) are the “uncertainty”, or “inaccuracy” in the simultaneous measurements of position and momentum for a single particle. Section 6.3 of [Nab5] discusses in some detail the physical origin of Heisenberg’s Uncertainty Principle, its current experimental status and the fact that it has nothing whatsoever to do with the inequality (A.10).
We will conclude our remarks on Postulate QM3 with a particularly significant special case. We let \( \psi \) denote a unit vector in \( \mathcal{H} \) representing some state. Now let \( \phi \) be another unit vector in \( \mathcal{H} \) representing another state. Then \( | \langle \psi, \phi \rangle |^2 \) is interpreted as the probability of finding the system in state \( \phi \) if it is known to be in state \( \psi \) before a measurement to determine the state is made; it is called the transition probability from state \( \psi \) to state \( \phi \). The complex number \( \langle \psi, \phi \rangle \) is called the transition amplitude from \( \psi \) to \( \phi \).

**Postulate QM4**

Let \( \mathcal{H} \) be the Hilbert space of an isolated quantum system. Then there exists a strongly continuous 1-parameter group \( \{ U_t \}_{t \in \mathbb{R}} \) of unitary operators on \( \mathcal{H} \), called evolution operators, with the property that, if the state of the system at time \( t = 0 \) is \( \psi_0 = \psi(0) \), then the state at time \( t \) is given by

\[
\psi_t = U_t(\psi_0) = U_t(\psi(0))
\]

for every \( t \in \mathbb{R} \). By Stone’s Theorem (Theorem 5.5.10 of [Nab5]) there is a unique self-adjoint operator \( H : \mathcal{D}(H) \to \mathcal{H} \), called the Hamiltonian of the system, such that \( U_t = e^{-itH/\hbar} \), where \( \hbar = \frac{\hbar}{2\pi} \) is the reduced Planck constant. Therefore

\[
\psi_t = \psi(t) = e^{-itH/\hbar}(\psi_0) = e^{-itH/\hbar}(\psi(0)) = e^{-itH/\hbar}(\psi(0)).
\]

Physically, the Hamiltonian \( H \) is identified with the operator representing the total energy of the quantum system.

We will view a quantum system as isolated if, as in the classical case, no matter or energy can enter or leave the system and if, in addition, no measurements are made on the system. One should be aware, however, that it is not at all clear that such things exist, nor is it clear precisely what we mean by a measurement. Rather than try to define what a measurement is we will adopt Postulate QM4 as a definition of what it means to say that measurements are not being performed on the system. We should point out also that the \( \hbar \) is introduced here simply to keep the units consistent with the interpretation of \( H \) as an energy (Remark 6.2.11 of [Nab5]).

The 1-parameter group \( \{ U_t \}_{t \in \mathbb{R}} \) of unitary operators is the analogue for an isolated quantum system of the classical 1-parameter group \( \{ \phi_t \}_{t \in \mathbb{R}} \) of diffeomorphisms describing the flow of the Hamiltonian vector field in classical mechanics; both “push” an initial state of the system onto the state at some future time. That the appropriate analogue of the diffeomorphism \( \phi_t \) is a unitary operator \( U_t \) deserves some comment. On the surface, the motivation seems clear. A state of the system is represented by a \( \psi \in \mathcal{H} \) with \( \langle \psi, \psi \rangle = 1 \) so the same must be true of the evolved states \( \phi(t) \), that is, we must have \( \langle \psi(t), \psi(t) \rangle = 1 \) for all \( t \in \mathbb{R} \). Certainly, this will be the case if \( \phi(t) \) is obtained from \( \psi(0) \) by applying a unitary operator \( U \) since

\[
\langle U(\psi(0)), U(\psi(0)) \rangle = \langle \psi(0), \psi(0) \rangle = 1.
\]

Notice, however, that this is also true if \( U \)}
is anti-unitary since then \( \langle U(\psi(0)), U(\psi(0)) \rangle = \langle \psi(0), \psi(0) \rangle = 1 \). Physically, one would probably also wish to assume that the time evolution preserves all transition probabilities \( |\langle \psi, \phi \rangle|^2 \), but this is also the case for both unitary and anti-unitary operators. Since unitary and anti-unitary operators differ only by a factor of \( \pm i \), one might be tempted to conclude that one choice is as good as the other. Physically, however, matters are not quite so simple (see [Wig3]). Furthermore, it is not so clear that there might not be other possibilities as well, that is, maps of \( \mathcal{H} \) onto \( \mathcal{H} \) that preserve transition probabilities but do not arise from unitary or anti-unitary operators.

That, in fact, there are no other possibilities is a consequence of a highly nontrivial result of Wigner ([Wig2]) that we described in Section 1.3. For convenience, we will repeat the description here under our current notation. We identify the state space of our quantum system with the projectivization \( P(\mathcal{H}) \) of \( \mathcal{H} \). For any \( \psi, \phi \in P(\mathcal{H}) \) we define the transition probability \( \langle \psi, \phi \rangle \) from state \( \psi \) to state \( \phi \) by

\[
\langle \psi, \phi \rangle = \frac{\langle \psi, \phi \rangle^2}{||\psi||^2 ||\phi||^2}
\]

for any \( \psi \in \mathcal{P} \) and any \( \phi \in \mathcal{F} \). Wigner defined a symmetry of the quantum system whose Hilbert space is \( \mathcal{H} \) to be a bijection \( T : \mathcal{P}(\mathcal{H}) \to \mathcal{P}(\mathcal{H}) \) that preserves transition probabilities in the sense that \( \langle T(\psi), T(\phi) \rangle = \langle \psi, \phi \rangle \) for all \( \psi, \phi \in \mathcal{P}(\mathcal{H}) \). Notice that, although \( \mathcal{P}(\mathcal{H}) \) has a natural quotient topology, no continuity assumptions are made. Any unitary or anti-unitary operator \( U \) on \( \mathcal{H} \) induces a symmetry \( T_U \) that carries any representative \( \psi \) of \( \mathcal{P} \) to the representative \( U(\psi) \) of \( T_U(\mathcal{P}) \). What Wigner proved was that every symmetry is induced in this way by a unitary or anti-unitary operator. The result is, in fact, a bit more general. There is a detailed proof of the following result in [Barg].

**Theorem A.4.2.** (Wigner’s Theorem on Symmetries) Let \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) be complex, separable Hilbert spaces and \( T : \mathcal{P}(\mathcal{H}_1) \to \mathcal{P}(\mathcal{H}_2) \) a mapping of \( \mathcal{P}(\mathcal{H}_1) \) into \( \mathcal{P}(\mathcal{H}_2) \) satisfying

\[
\langle T(\psi), T(\phi) \rangle_{\mathcal{P}(\mathcal{H}_2)} = \langle \psi, \phi \rangle_{\mathcal{P}(\mathcal{H}_1)}
\]

for all \( \psi, \phi \in \mathcal{P}(\mathcal{H}_1) \). Then there exists a mapping \( U : \mathcal{H}_1 \to \mathcal{H}_2 \) satisfying

1. \( U(\psi) \in T(\psi) \) if \( \psi \in \mathcal{P}(\mathcal{H}_1) \),

2. \( U(\psi + \phi) = U(\psi) + U(\phi) \) for all \( \psi, \phi \in \mathcal{H}_1 \), and

3. either
   a. \( U(\lambda \psi) = \lambda U(\psi) \) and \( \langle U(\psi), U(\phi) \rangle_{\mathcal{H}_2} = \langle \psi, \phi \rangle_{\mathcal{H}_1} \), or
   b. \( U(\lambda \psi) = \bar{\lambda} U(\psi) \) and \( \langle U(\psi), U(\phi) \rangle_{\mathcal{H}_2} = \overline{\langle \psi, \phi \rangle_{\mathcal{H}_1}} \)

for all \( \lambda \in \mathbb{C} \) and all \( \psi, \phi \in \mathcal{H}_1 \).
Furthermore, if \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) are of dimension at least 2 and \( T((\psi_1)) = T((\psi_2)) \), where \( \psi_1 \) and \( \psi_2 \) are unit vectors in \( \mathcal{H}_1 \), then there is a unique such mapping \( U : \mathcal{H}_1 \to \mathcal{H}_2 \) for which \( U(\psi_1) = U(\psi_2) \).

In particular, any symmetry of a quantum system with Hilbert space \( \mathcal{H} \) arises from an operator on \( \mathcal{H} \) that is either unitary or anti-unitary. The anti-unitary operators correspond physically to discrete symmetries such as time reversal (they are the analogue of reflections in Euclidean geometry).

Now we can justify the unitarity assumption in Postulate QM4. Suppose that the time evolution is described by an assignment to each \( t \in \mathbb{R} \) of a symmetry \( \alpha_t : \mathcal{P} \mathcal{H} \to \mathcal{P} \mathcal{H} \) and that \( t \to \alpha_t \) satisfies \( \alpha_{t+s} = \alpha_t \circ \alpha_s \) for all \( t, s \in \mathbb{R} \). Then, for any \( t \in \mathbb{R} \), \( \alpha_t = \alpha_{t/2} \). By Wigner’s Theorem, \( \alpha_{t/2} \) is represented by an operator \( U_{t/2} \) that is either unitary or anti-unitary. Since the square of an operator that is either unitary or anti-unitary is necessarily unitary, every \( U_t \) must be unitary.

**Remark A.4.2.** We investigate the consequences of Wigner’s Theorem more thoroughly in Section 1.3, but a few remarks here would seem to be in order. A bijection \( T : \mathcal{P} \mathcal{H} \to \mathcal{P} \mathcal{H} \) preserving transition probabilities that is a homeomorphism with respect to the quotient topology is called an automorphism of \( \mathcal{P} \mathcal{H} \) and the collection of all such is clearly a group under composition. This is called the automorphism group of \( \mathcal{P} \mathcal{H} \) and is denoted \( \text{Aut}(\mathcal{P} \mathcal{H}) \). Wigner’s Theorem will permit us, in Section 1.3, to provide \( \text{Aut}(\mathcal{P} \mathcal{H}) \) with a natural topology with respect to which it is a Hausdorff topological group. Now suppose that \( G \) is a Lie group. A continuous homomorphism of \( G \) into \( \text{Aut}(\mathcal{P} \mathcal{H}) \) is called a projective representation of \( G \) on \( \mathcal{H} \). If \( \mathcal{H} \) is the Hilbert space of some quantum system, then such a projective representation assigns a symmetry to each element of \( G \) in such a way that the group operations are respected and this leads us to refer to \( G \) as a symmetry group of the quantum system. When \( G \) is the Poincaré group \( \mathcal{P} \mathcal{L} \), the existence of such a projective representation is the natural expression of the “relativistic invariance” of the quantum system.

Notice that there is nothing special about \( t = 0 \) in Postulate QM4. If \( t_0 \) is any real number, then \( \psi(t_0) = U_{t_0}(\psi(0)) \) so \( \psi(t + t_0) = U_{t_0}(\psi(t)) = U_{t_0}(U_{t_0}(\psi(0))) \) and therefore

\[
\psi(t) = \psi((t - t_0) + t_0) = U_{t_0}(\psi(t_0)) = e^{-itH\hbar}(\psi(t_0)).
\]

Thus,

\[
U_{t_0} = e^{-itH\hbar}
\]

propagates the state at time \( t_0 \) to the state at time \( t \) for any \( t_0, t \in \mathbb{R} \). It follows from this that if \( \psi(t) \) is thought of as a curve in \( \mathcal{H} \) and \( \psi(t_0) \) is in the domain of \( H \), then, by Stone’s Theorem, \( \psi(t) \) is in the domain of \( H \) for all \( t \in \mathbb{R} \) and
\[ i\hbar \frac{d\psi(t)}{dt} = H(\psi(t)), \]  
\[ (A.12) \]

where the derivative is defined by

\[ \frac{d\psi(t)}{dt} = \lim_{t \to t_0} \frac{\psi(t) - \psi(t_0)}{t - t_0} \]  
\[ (A.13) \]

and the limit is in $\mathcal{H}$. Equation (A.12) is called the *abstract Schrödinger equation*.

This, however, is not the way one generally sees the Schrödinger equation written in the physics literature. When $H$ is, for example, $L^2(\mathbb{R}^n)$, then each state of the system is represented by a complex-valued, square integrable wave function $\psi(q)$ on $\mathbb{R}^n$ with $L^2$-norm 1. In this case the Hamiltonian $H$ is generally a differential operator of the form

\[ H = H_0 + V(q) = -\frac{\hbar^2}{2m} \Delta + V(q), \]  
\[ (A.14) \]

where $V(q)$ is a real-valued function on $\mathbb{R}^n$, called the *potential*, which acts on $L^2(\mathbb{R}^n)$ as a multiplication operator, $m$ is a positive constant, and $\Delta$ is the distributional Laplacian on $L^2(\mathbb{R}^n)$.

**Remark A.4.3.** The operators $H_0$ and $V$ are both self-adjoint on $L^2(\mathbb{R}^n)$, but generally their sum is not. $V$ must satisfy conditions that are sufficient to ensure that the operator $H$ is self-adjoint on $L^2(\mathbb{R}^n)$ and finding such conditions is not at all trivial. There is a discussion of some results of this sort in Section 8.4.2 of [Nab5].

The time evolution of the wave function is then generally written $\psi(q, t)$ rather than $\psi(t)$ or $(\psi(t))(q)$ and the Schrödinger equation is written

\[ i\hbar \frac{\partial \psi(q, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi(q, t) + V(q)\psi(q, t). \]  
\[ (A.15) \]

The point we would like to stress, however, is that writing the Schrödinger equation in the form (A.15) involves a fair amount of potentially hazardous notational subterfuge. The abstract Schrödinger equation (A.12) contains in it the $t$-derivative $d\psi(t)/dt$ which is defined as the limit in $L^2(\mathbb{R}^n)$ of the familiar difference quotient (see (A.13)). The partial derivative $\partial \psi(q, t)/\partial t$ that appears in the traditional physicist’s form (A.15) of the Schrödinger equation is, on the other hand, defined as a limit in $C$ of an equally familiar difference quotient. In general, there is no reason to suppose that the latter exists for $\psi \in L^2(\mathbb{R}^n)$ and, even if it does, that it is in $L^2(\mathbb{R}^n)$ for each $t$ and, granting even this, that it is equal to $d\psi(t)/dt$. Sufficient regularity assumptions on $\psi(q, t)$ will guarantee that all of these things are true (Exercise 6.2.1 of [Nab5]), but even these may not be enough to ensure that, for each $t$, the distribution $\Delta \psi$ is actually a function in $L^2(\mathbb{R}^n)$ and, if this is not the case, we have no business equating it in (A.15) to something that is a function in $L^2(\mathbb{R}^n)$. All of these difficulties can be willed away by restricting attention to functions $\psi(q, t)$ that are
sufficiently regular, say, continuously differentiable with respect to \( t \) and, for each \( t \), a Schwartz function or smooth function with compact support on \( \mathbb{R}^n \). The hope then is that one can find sufficiently many such classical solutions to (A.15) to provide an orthonormal basis for \( L^2(\mathbb{R}^n) \) for each \( t \). Example 5.3.1 of [Nab5] shows that, at least in the case of the harmonic oscillator, one’s hopes are not dashed.

We will conclude this section with a brief synopsis the salient features of the picture of quantum mechanics we have painted thus far and then suggest a rather different way of looking at it that will bear a striking resemblance to the picture of classical Hamiltonian mechanics in Section A.3. A quantum system has associated with it a complex, separable Hilbert space \( \mathcal{H} \) and a distinguished self-adjoint operator \( H \), called the Hamiltonian of the system. The states of the system are represented by unit vectors \( U_i \) in \( \mathcal{H} \) and these evolve in time from an initial state \( \psi(0) \) according to

\[
\psi(t) = U_t(\psi(0)) = e^{-iHt/\hbar}(\psi(0)).
\]

As a result, the evolving states satisfy the abstract Schrödinger equation

\[
i\hbar \frac{d\psi(t)}{dt} = H(\psi(t)).
\]

(A.16)

Each observable is identified with a self-adjoint operator \( A \) that does not change with time. Neither the state vectors \( \psi \) nor the observables \( A \) are accessible to direct experimental measurement. Rather, the link between the formalism and the physics is contained in the expectation values \( \langle A \rangle_\psi = \langle \psi, A\psi \rangle \). Knowing these one can construct the probability measures \( \mu_{A,\psi}(S) = \langle \psi, E^A(S)\psi \rangle \) and these contain all of the information that quantum mechanics permits us to know about the system.

We would now like to look at this from a slightly different point of view. As the state evolves so do the expectation values of any observable. Specifically,

\[
\langle A \rangle_\psi(t) = \langle \psi(t), A\psi(t) \rangle = \langle U_t(\psi(0)), A U_t(\psi(0)) \rangle = \langle \psi(0), [U_t^{-1}A U_t] \psi(0) \rangle,
\]

because each \( U_t \) is unitary. Now, define a (necessarily self-adjoint) operator

\[
A(t) = U_t^{-1} A U_t
\]

for each \( t \in \mathbb{R} \). Then

\[
\langle A \rangle_{\psi(t)} = \langle A(t) \rangle_{\psi(0)}
\]

for each \( t \in \mathbb{R} \). The expectation value of \( A \) in the evolved state \( \psi(t) \) is the same as the expectation value of the observable \( A(t) \) in the initial state \( \psi(0) \). Since all of the physics is contained in the expectation values this presents us with the option of regarding the states as fixed and the observables as evolving in time. From this point of view our quantum system has a fixed state \( \psi \) and the observables evolve in time from some initial self-adjoint operator \( A = A(0) \) according to

\[
A(t) = U_t^{-1} A U_t = e^{iHt/\hbar} A e^{-iHt/\hbar}.
\]

(A.17)
This is called the **Heisenberg picture** of quantum mechanics to distinguish it from the view we have taken up to this point, which is called the **Schrödinger picture**. Although these two points of view appear to differ from each other rather trivially, the Heisenberg picture occasionally presents some significant advantages and we will now spend a moment seeing what things look like in this picture.

We should first notice that, when $A$ is the Hamiltonian $H$ itself, Stone’s Theorem implies that each $U_t$ leaves $\mathcal{D}(H)$ invariant and commutes with $H$ so

$$H(t) = U_t^{-1}HU_t = e^{itH/h}He^{-itH/h} = H \quad \forall t \in \mathbb{R}.$$ 

The Hamiltonian is constant in time in the Heisenberg picture. For other observables this is generally not the case, of course, and one would like to have a differential equation describing their time evolution in the same way that the Schrödinger equation describes the time evolution of the states in the Schrödinger picture. We will describe such an equation in the case of observables represented by **bounded** self-adjoint operators in the Schrödinger picture.

**Remark A.4.4.** This is a very special case, of course, so we should explain the restriction. In the unbounded case, a rigorous derivation of the equation is substantially complicated by the fact that, in the Heisenberg picture, the operators (and therefore their domains), are varying with $t$ so that the usual domain issues for unbounded operators also vary with $t$. Physicists have the good sense to ignore all of these issues and just formally differentiate (A.17), thereby arriving at the very same equation that appears in our theorem below. Furthermore, it is not hard to show that, if $A$ is unbounded and $A(t) = U_t^{-1}AU_t$, then, for any Borel function $f$, $f(A(t)) = f(A)(t) = U_t^{-1}f(A)U_t$ so that one can generally study the time evolution of $A$ in terms of the time evolution of the bounded functions of $A$ and these are bounded operators. In particular we recall that, from the point of view of physics, all of the relevant information is contained in the probability measures $\langle \psi, E_A(S)\psi \rangle$ so that, in principle, one requires only the time evolution of the (bounded) projections $E_A(S)$.

The following is Theorem 6.4.1 of [Nab5].

**Theorem A.4.3.** Let $\mathcal{H}$ be a complex, separable Hilbert space, $H : \mathcal{D}(H) \to \mathcal{H}$ a self-adjoint operator on $\mathcal{H}$ and $U_t = e^{-itH/h}$, $t \in \mathbb{R}$, the 1-parameter group of unitary operators determined by $H$. Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded, self-adjoint operator on $\mathcal{H}$ and define, for each $t \in \mathbb{R}$, $A(t) = U_t^{-1}AU_t$. If $\psi$ and $A(t)\psi$ are in $\mathcal{D}(H)$ for every $t \in \mathbb{R}$, then $A(t)\psi$ satisfies the **Heisenberg equation**

$$\frac{dA(t)}{dt} \psi = -\frac{i}{\hbar} [A(t), H] \psi, \quad (A.18)$$

where the derivative is the $\mathcal{H}$-limit.
\[ \frac{dA(t)}{dt} \psi = \lim_{\Delta t \to 0} \left[ \frac{A(t + \Delta t) - A(t)}{\Delta t} \right] \psi \]

and \([A(t), H] \) is the commutator of \(A(t)\) and \(H\) on \(D(H)\), that is, \([A(t), H] \psi = A(t)(H\psi) - H(A(t)\psi)\) for every \(\psi \in D(H)\).

Let’s simplify the notation a bit and write (A.18) as

\[ \frac{dA}{dt} = -i \frac{\hbar}{2} [A, H]. \]  

(A.19)

Now compare this with the equation (A.9)

\[ \frac{df}{dt} = \{f, H\} \]

describing the time evolution of a classical observable in the Hamiltonian picture of mechanics. The analogy is striking and suggested to Paul Dirac [Dirac1] a possible avenue from classical to quantum mechanics, that is, a possible approach to the quantization of classical mechanical systems. The idea is that classical observables should be replaced by self-adjoint operators and the Poisson bracket \(\{, \}\) by the quantum bracket

\[ \{, \}_h = -i \frac{\hbar}{2} [\{, \}]. \]

Let’s spell this out in a bit more detail. Dirac’s suggestion was to find a linear map from the classical observables \(f, g, \ldots\) to the quantum observables \(F, G, \ldots\) with the property that

\[ \{f, g\} \mapsto [F, G]_h = -i \frac{\hbar}{2} [F, G]. \]

If one further stipulates that the constant function 1 should map to the identity operator \(I\) then this implies that the classical canonical commutation relations (A.8)

\[ \{q^i, q^j\} = \{p_i, p_j\} = 0 \quad \text{and} \quad \{q^i, p_j\} = \delta^i_j, \quad i, j = 1, \ldots, n \]  

(A.20)

map to the quantum canonical commutation relations

\[ \{Q^i, Q^j\}_h = \{P_i, P_j\}_h = 0 \quad \text{and} \quad \{Q^i, P_j\}_h = \delta^i_j I, \quad i, j = 1, \ldots, n, \]  

(A.21)

where \(Q^i\) and \(P_i\) are the images of \(q^i\) and \(p_i\), respectively, for \(i = 1, \ldots, n\). Written in terms of commutators, (A.21) becomes

\[ [Q^i, Q^j] = [P_i, P_j] = 0 \quad \text{and} \quad [Q^i, P_j] = i\hbar \delta^i_j I, \quad i, j = 1, \ldots, n. \]  

(A.22)

Presumably the image under this map of the classical Hamiltonian would be the appropriate quantum Hamiltonian and with this in hand the analysis of the quan-
tum system could commence. The extent to which Dirac’s program can actually be carried out is discussed in some detail in Section 7.2 of [Nab5].

Example A.4.1. We will conclude with a brief synopsis of the quantum system describing a single particle of mass $m$ moving in $\mathbb{R}^3$ under the influence of a time-independent potential $V(q) = V(q^1, q^2, q^3)$. The corresponding problem for motion in $\mathbb{R}$ is treated in considerable detail in [Nab5] and we will provide references for those results whose extension from one to three spatial dimensions is not routine.

Remark A.4.5. We should point out that, in this example, we will be ignoring an important quantum mechanical property of elementary particles called spin. We will take up this subject and see what modifications of our present discussion are required in Section A.5.

The Hilbert space $\mathcal{H}$ of our quantum system is taken to be $L^2(\mathbb{R}^3)$, that is, the Hilbert space of (equivalence classes of) complex-valued, square integrable functions on $\mathbb{R}^3$ with respect to Lebesgue measure. The dynamics of the system is governed, via the Schrödinger equation, by the Hamiltonian $H$ which must be a self-adjoint operator on $L^2(\mathbb{R}^3)$ representing the total energy of the system. The total energy is the sum of the kinetic energy and the potential energy so $H$ will be the sum of two self-adjoint operators. The kinetic term is always the same and is defined in the following way. Begin by looking at the subspace $C^\infty_0(\mathbb{R}^3)$ of $L^2(\mathbb{R}^3)$ consisting of smooth, complex-valued functions with compact support on $\mathbb{R}^3$ (or the Schwartz space $S(\mathbb{R}^3)$ of rapidly decreasing complex-valued functions on $\mathbb{R}^3$). On this subspace the ordinary Laplacian $\Delta$ is defined and essentially self-adjoint (see pages 395-396 of [Nab5]) so the same is true of

$$H_0 = -\frac{\hbar^2}{2m} \Delta.$$

The minus sign ensures that $H_0$ is a positive operator, that is,

$$\langle H_0 \psi, \psi \rangle \geq 0$$

for all $\psi \in C^\infty_0(\mathbb{R}^3)$ (or $S(\mathbb{R}^3)$). The unique self-adjoint extension is denoted with the same symbols and its domain is

$$\mathcal{D}(H_0) = \{ \psi \in L^2(\mathbb{R}^3) : A\psi \in L^2(\mathbb{R}^3) \},$$

where $A$ now means the distributional Laplacian defined by taking Fourier transforms (see Sections 8.4.1 and 8.4.2 of [Nab5]). $H_0$ is the kinetic energy term in our Hamiltonian.

Remark A.4.6. The motivation for adopting $H_0$ as the kinetic energy operator is that it is the canonical quantization of the classical kinetic energy (see Chapter 7 of [Nab5]).
The potential $V$ is assumed to be a real-valued measurable function on $\mathbb{R}^3$ and, as an operator on $L^2(\mathbb{R}^3)$, it acts by multiplication $(V\psi)(q) = V(q)\psi(q))$. It is defined and essentially self-adjoint on $C^\infty_0(\mathbb{R}^3)$ (or $S(\mathbb{R}^3)$) and its unique self-adjoint extension, still denoted $V$, has domain

$$D(V) = \{\psi \in L^2(\mathbb{R}^3) : V\psi \in L^2(\mathbb{R}^3)\}.$$

This is the potential energy term in the Hamiltonian. One would now like to define the Hamiltonian $H$ by $H = H_0 + V$. However, $H_0$ and $V$ are both unbounded operators on $L^2(\mathbb{R}^3)$ and there is no a priori reason to suppose that their domains intersect in anything more than $0 \in L^2(\mathbb{R}^3)$. Moreover, even if the intersection of their domains happens to be a dense linear subspace, it is not true that the sum of two unbounded, self-adjoint operators is self-adjoint. In order to guarantee self-adjointness one must impose additional restrictions on $V$. Many such conditions are known and some of these are described in Section 8.4.2 of [Nab5]. We will state only the one that is actually proved in [Nab5] and refer those interested in seeing more to Chapter X of [RS2].

**Theorem A.4.4.** Let $V$ be a real-valued, measurable function on $\mathbb{R}^3$ that can be written as $V = V_1 + V_2$, where $V_1 \in L^2(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$ and $V_2 \in L^\infty(\mathbb{R}^3)$. Then

$$H = H_0 + V = -\frac{\hbar^2}{2m}A + V$$

is essentially self-adjoint on $C^\infty_0(\mathbb{R}^3)$ and self-adjoint on $D(H_0)$.

With the Hilbert space $L^2(\mathbb{R}^3)$ and a self-adjoint Hamiltonian $H$ in hand one next introduces self-adjoint operators that will act as the observables of the system. Initially, at least, one would look for observables that can be regarded as quantum analogues of the appropriate observables for the corresponding classical system. In the case at hand (a particle of mass $m$ moving in $\mathbb{R}^3$) these would include the position, energy, momentum and angular momentum. The energy, of course, is just the Hamiltonian $H$. Choosing appropriate operators to represent position, momentum, angular momentum or any other observable of interest is called quantization and this is not a process that admits a simple algorithmic synopsis. Those interested in a brief look behind the scenes at what is involved may want to refer to Chapter 7 of [Nab5]. Here we will simply record the operators of interest to us at the moment.

We fix an orthonormal basis for $\mathbb{R}^3$ and denote the corresponding coordinates by $q_1, q_2, q_3$. For each $j = 1, 2, 3$ we define the $j$th-coordinate position operator $Q^j$ on $L^2(\mathbb{R}^3)$ by

$$(Q^j\psi)(q) = (Q^j\psi)(q^1, q^2, q^3) = q^j\psi(q^1, q^2, q^3) \tag{A.23}$$

(motivation for the corresponding definition in one spatial dimension is provided in Remark 6.2.8 of [Nab5]). Then $Q^j$ is defined and self-adjoint on
\[ \mathcal{D}(Q^j) = \{ \psi \in L^2(\mathbb{R}^3) : q^i \psi(q^1, q^2, q^3) \in L^2(\mathbb{R}^3) \}. \]  

(A.24)

Example 5.2.6 of [Nab5] gives two proofs of the corresponding statement for motion in one spatial dimension that can easily be adapted to the 3-dimensional context in which we currently find ourselves; alternatively, Section 4.5, Chapter III, of [Prug] provides a different argument that proves self-adjointness in any number of spatial dimensions.

Next we would like to define, for each \( j = 1, 2, 3 \), the \( j \)-th coordinate momentum operator \( P_j \) on \( L^2(\mathbb{R}^3) \). One begins by showing that the operator

\[ P_j = -i\hbar \frac{\partial}{\partial q^j} \]  

(A.25)

is defined and essentially self-adjoint on \( C_0^{\infty}(\mathbb{R}^3) \) (or \( S(\mathbb{R}^3) \)) and taking the momentum operator, denoted with the same symbol, to be its unique self-adjoint extension (motivation for the corresponding definition in one spatial dimension is provided in Remark 6.2.15 of [Nab5]). The domain \( \mathcal{D}(P_j) \) is best viewed in the following way. The Fourier transform \( \mathcal{F} : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3) \) is a unitary operator on \( L^2(\mathbb{R}^3) \). For \( \psi \) in \( C_0^{\infty}(\mathbb{R}^3) \) (or \( S(\mathbb{R}^3) \)) one has \( P_j \psi = (\mathcal{F}^{-1}(\hbar Q^j)\mathcal{F})\psi \). Consequently,

\[ P_j = \mathcal{F}^{-1}(\hbar Q^j)\mathcal{F} \]  

(A.26)

on \( C_0^{\infty}(\mathbb{R}^3) \) (or \( S(\mathbb{R}^3) \)). Since each of these is dense in \( L^2(\mathbb{R}^3) \), (A.26) is true everywhere on \( L^2(\mathbb{R}^3) \). In particular, \( P_j \) is unitarily equivalent to \( Q^j \) and the domain of \( P_j \) is

\[ \mathcal{D}(P_j) = \mathcal{F}^{-1}(\mathcal{D}(Q^j)). \]  

(A.27)

One often sees (A.26) taken as the definition of \( P_j \) in which case the self-adjointness of \( P_j \) follows from its unitary equivalence to \( Q^j \) (Lemma 5.2.5 of [Nab5]).

Finally, we would like to write out the operators on \( L^2(\mathbb{R}^3) \) that represent the quantum analogues of the classical components of orbital angular momentum. The motivation is to be found in the infinitesimal generators (A.3), (A.4) and (A.5) of classical angular momentum. Specifically, we begin by defining operators \( M^{23} \), \( M^{31} \) and \( M^{12} \) on \( C_0^{\infty}(\mathbb{R}^3) \) (or \( S(\mathbb{R}^3) \)) by

\[ M^{23} = Q^2 P^3 - Q^3 P^2 = i \hbar \left( q^3 \frac{\partial}{\partial q^2} - q^2 \frac{\partial}{\partial q^3} \right) \]  

(A.28)

\[ M^{31} = Q^3 P^1 - Q^1 P^3 = i \hbar \left( q^1 \frac{\partial}{\partial q^3} - q^3 \frac{\partial}{\partial q^1} \right) \]  

(A.29)

\[ M^{12} = Q^1 P^2 - Q^2 P^1 = i \hbar \left( q^2 \frac{\partial}{\partial q^1} - q^1 \frac{\partial}{\partial q^2} \right) \]  

(A.30)
where \( P^j = \eta^{jk} P_k = -P_j = i\hbar \frac{\partial}{\partial q_j} \) for \( j = 1, 2, 3 \).

**Remark A.4.7.** See Example 2.5.1 and what follows for more on these operators. Note, however, that there we have taken \( \hbar = 1 \).

The operators \( M^{23}, M^{31} \) and \( M^{12} \) are essentially self-adjoint on \( C_0^\infty(\mathbb{R}^3) \) (or \( S(\mathbb{R}^3) \)) and their unique self-adjoint extensions, denoted with the same symbols, are the components of the orbital angular momentum operators in quantum mechanics.

Now recall that the Fourier transform \( \mathcal{F} \) on \( \mathbb{R}^n \) satisfies each of the following for any \( \hat{\psi} \in \mathcal{S}(\mathbb{R}^n) \) and any \( j = 1, \ldots, n \).

\[
\mathcal{F}\left( \frac{\partial}{\partial q_j} \psi \right)(p) = i p_j \mathcal{F}(\psi)(p)
\]

\[
\mathcal{F}\left( -i q_j \psi \right)(p) = \frac{\partial}{\partial p_j} \mathcal{F}(\psi)(p)
\]

and is a unitary operator of \( L^2(\mathbb{R}^n, d^n q) \) onto \( L^2(\mathbb{R}^n, d^n p) \). Any operator \( A \) on \( L^2(\mathbb{R}^n, d^n q) \) can therefore be regarded as an operator on \( L^2(\mathbb{R}^n, d^n p) \), specifically, the operator \( \mathcal{F} A \mathcal{F}^{-1} \).

**Exercise A.4.1.** Write the Fourier transform \( \mathcal{F}(\psi) \) of \( \hat{\psi} \) and prove each of the following.

1. \( \left( [\mathcal{F} \hat{Q} \mathcal{F}^{-1}] \hat{\psi} \right)(p) = i \frac{\partial}{\partial p_j} \hat{\psi}(p), \quad j = 1, 2, 3 \)
2. \( \left( [\mathcal{F} \hat{P}_k \mathcal{F}^{-1}] \hat{\psi} \right)(p) = \hbar p_k \hat{\psi}(p), \quad k = 1, 2, 3 \)
3. \( \left( [\mathcal{F} \hat{Q} \mathcal{F}^{-1}] [\mathcal{F} \hat{P}_k \mathcal{F}^{-1}] \hat{\psi} \right)(p) = -i \hbar p_k \frac{\partial}{\partial p_j} \hat{\psi}(p), \quad j, k = 1, 2, 3, \quad j \neq k \)

**Exercise A.4.2.** Show that, as operators on momentum space \( L^2(\mathbb{R}^3, d^3 p) \), the operators \( M^{23}, M^{31}, \) and \( M^{12} \) take the same form as on \( L^2(\mathbb{R}^3, d^3 q) \), that is,

\[
M^{23} = i \hbar \left( \frac{1}{p_1} \frac{\partial}{\partial p_2} - \frac{1}{p_2} \frac{\partial}{\partial p_3} \right), \quad (A.31)
\]

\[
M^{31} = i \hbar \left( \frac{1}{p_1} \frac{\partial}{\partial p_3} - \frac{1}{p_3} \frac{\partial}{\partial p_1} \right), \quad (A.32)
\]

\[
M^{12} = i \hbar \left( \frac{1}{p_2} \frac{\partial}{\partial p_1} - \frac{1}{p_1} \frac{\partial}{\partial p_2} \right). \quad (A.33)
\]

As we mentioned in Remark A.4.5 the discussion of particle motion in quantum mechanics in Example A.4.1 took no account of the quantum mechanical property of *spin*; more precisely, the discussion is valid only for particles of *spin zero* and we do not yet know what this means. In the next section we will try to remedy this (more information is available in Chapter 9 of [Nab5]).
A.5 Spin

We will try to provide some sense of what this phenomenon of spin is and how the physicists have incorporated it into their mathematical model of the quantum world. There is nothing like quantum mechanical spin in classical physics. There is, however, a classical analogy. The analogy is inadequate and can be misleading if taken too seriously, but it is the best we can do so we will begin by briefly describing it.

Imagine a spherical mass $m$ of radius $a$ moving through space on a circular orbit of radius $R \gg a$ about some point $O$ and, at the same time, spinning around an axis through one of its diameters (to a reasonable approximation, the Earth does all of this). Due to its orbital motion, the mass has an angular momentum $L = r \times (mv) = r \times p$, where $r$ is the position vector from $O$ and $v = \dot{r}$ is the velocity, which we call its orbital angular momentum. The spinning of the mass around its axis contributes additional angular momentum that one calculates by subdividing the spherical region occupied by the mass into subregions, regarding each subregion as a mass in a circular orbit about a point on the axis, approximating its angular momentum, adding all of these and taking the limit as the regions shrink to points. The resulting integral gives the angular momentum due to rotation. This is called the rotational angular momentum, is denoted $S$, and is given by

$$S = I\omega,$$

where $I$ is the moment of inertia of the sphere and $\omega$ is the angular velocity ($\omega$ is along the axis of rotation in the direction determined by the right-hand rule from the direction of the rotation). If the mass is assumed to be uniformly distributed throughout the sphere (in other words, if the sphere has constant density), then an exercise in calculus gives

$$S = \frac{2}{5}ma^2\omega.$$ 

The total angular momentum of the sphere is $L + S$.

Now let’s suppose, in addition, that the sphere is charged. Due to its orbital motion the charged sphere behaves like a current loop and any moving charge gives rise to a magnetic field. If we assume that our current loop is very small (or, equivalently, that we are viewing it from a great distance) the corresponding magnetic field is that of a magnetic dipole (see Sections 14-5 and 34-2, Volume II, of [FLS]). All we need to know about this is that this magnetic dipole is described by a vector $\mu_L$ called its orbital magnetic moment that is proportional to the orbital angular momentum. Specifically,

$$\mu_L = \frac{q}{2m}L,$$

where $q$ is the charge of the sphere (which can be positive or negative). Similarly, the rotational angular momentum of the charge gives rise to a magnetic field that is
also that of a magnetic dipole and is described by a rotational magnetic moment $\mu_S$ given by

$$\mu_S = \frac{q}{2m} S.$$ 

The total magnetic moment $\mu$ is

$$\mu = \mu_L + \mu_S = \frac{q}{2m} (L + S).$$

The significance of the magnetic moment $\mu$ of the dipole is that it describes the strength and direction of the dipole field and determines the torque

$$\tau = \mu \times B$$

experienced by the magnetic dipole when placed in an external magnetic field $B$. If the magnetic field $B$ is uniform (that is, constant), then its only effect on the dipole is to force $\mu$ to precess around a cone whose axis is along $B$ in the same way that the axis of a spinning top precesses around the direction of the Earth’s gravitational field (see Figure A.1 and Section 2, Chapter 11, of [Eis]). Notice that this precession does not change the projection $\mu \cdot B$ of $\mu$ along $B$.

![Precession](image)

Fig. A.1 Precession

If the $B$-field is not uniform, however, there will be an additional translational force acting on the mass which, if $m$ is moving through the field, will push it off the course it would have followed if $B$ had been uniform. Precisely what this deflection will be depends, of course, on the nature of $B$ and we will say a bit more about this in a moment.

Now we will describe the famous Stern-Gerlach experiment (a schematic of which is shown in Figure A.2). We are interested in whether or not the electron has a rotational magnetic moment and, if so, whether or not its behavior is ade-
quately described by classical physics. What we will do is send a certain beam of electrically neutral atoms through a non-uniform magnetic field $\mathbf{B}$ and then let them hit a photographic plate to record how their paths were deflected by the field. The atoms must be electrically neutral so that the deflections due to the charge do not mask any deflections due to magnetic moments of the atoms. In particular, we can’t do this with free electrons. The atoms must also have the property that any magnetic moment they might have could be due only to a single electron somewhere within it. Stern and Gerlach chose atoms of silver ($\text{Ag}$) which they obtained by evaporating the metal in a furnace and focusing the resulting gas of $\text{Ag}$ atoms into a beam aimed at a magnetic field.

Remark A.5.1. Silver is a good choice, but for reasons that are not so apparent. A proper explanation requires some hindsight (not all of the information was available to Stern and Gerlach) as well as some quantum mechanical properties of atoms that we have not discussed here. Nevertheless, it is worth saying at least once since otherwise one is left with all sorts unanswered questions about the validity of the experiment. So, here it is. The stable isotopes of $\text{Ag}$ have 47 electrons, 47 protons and either 60 or 62 neutrons so, in particular, they are electrically neutral. Since the magnetic moment is inversely proportional to the mass and since the mass of the proton and neutron are each approximately 2000 times the mass of the electron, one can assume that any magnetic moments of the nucleons will have a negligible effect on the magnetic moment of the atom and can therefore be ignored. Of the 47 electrons, 46 are contained in contained in closed, inner shells (energy levels) and these, it turns out, can be represented as a spherically symmetric cloud with no orbital or
rotational angular momentum (this is not at all obvious). The remaining electron is in what is termed the outer 5s-shell and an electron in an s-state has no orbital angular momentum (again, not obvious). Granting all of this, the only possible source of any magnetic moment for a Ag atom is a rotational angular momentum of its outer 5s-electron. Whatever happens in the experiment is attributable to the electron and the rest of the silver atom is just a package designed to ensure this.

We will first see what the classical picture of an electron with a rotational magnetic moment would lead us to expect in the Stern-Gerlach experiment and will then describe the results that Stern and Gerlach actually obtained (a more thorough, but quite readable account of the physics is available Chapter 11 of [Eis]). For this we will need to be more specific about the magnetic field $B$ that we intend to send the Ag atoms through. Let’s introduce a coordinate system in Figure A.2 in such a way that the Ag atoms move in the direction of the $y$-axis and the vertical axis of symmetry of the magnet is along the $z$-axis so that the $x$-axis is perpendicular to both of these. The magnet itself can be designed to produce a field that is non-uniform, but does not vary with $y$, is predominantly in the $z$-direction, and is symmetric with respect to the $yz$-plane. The interaction between the neutral Ag atom (with magnetic moment $\mu$) and the non-uniform magnetic field $B$ provides the atom with a potential energy $\mu \cdot B$ so that the atom experiences a force

$$ F = \nabla(\mu \cdot B) = \nabla(\mu_x B_x + \mu_y B_y + \mu_z B_z). $$

For the sort of magnetic field we have just described, $B_y = 0$ and $B_z$ dominates $B_x$. From this one finds that the translational motion is governed primarily by

$$ F_z \approx \mu_z \frac{\partial B_z}{\partial z} $$

(see pages 333-334 of [Eis]). The conclusion we draw from this is that the displacements from the intended path of the silver atoms will be in the $z$-direction (up and down in Figure A.2) and the forces causing these displacements are proportional to the $z$-component of the magnetic moment. Of course, different orientations of the magnetic moment $\mu$ among the various Ag atoms will lead to different values of $\mu_z$ and therefore to different displacements. Moreover, due to the random thermal effects of the furnace, one would expect that the silver atoms exit with their magnetic moments $\mu$ randomly oriented in space so that their $z$-components could take on any value in the interval $[-|\mu|, |\mu|]$. As a result, the expectation based on classical physics would be that the deflected Ag atoms will impact the photographic plate at points that cover an entire vertical line segment (see the segment labeled “Classical prediction” in Figure A.2).

**Remark A.5.2.** Writing $q = -e$ for the charge of the electron and $m = m_e$ for its mass we find that
$F_z \approx \mu_z \frac{\partial B_z}{\partial z} = -\frac{e}{2m_e} S_z \frac{\partial B_z}{\partial z}$

so that the deflection of an individual Ag atom is a measure of the component $S_z$ of $S$ in the direction of the magnetic field gradient.

This, however, is not at all what Stern and Gerlach observed. What they found was that the silver atoms arrived at the screen at only two points, one above and one the same distance below the y-axis (again, see Figure A.2). The experiment was repeated with different orientations of the magnet (that is, different choices for the z-axis) and different atoms and nothing changed. We seem to be dealing with a very peculiar sort of “vector” $S$. The classical picture would have us believe that, however it is oriented in space, its projection onto any axis is always one of two things. Needless to say, ordinary vectors in $\mathbb{R}^3$ do not behave this way. What we are really being told is that the classical picture is simply wrong. The property of electrons that manifests itself in the Stern-Gerlach experiment is in some ways analogous to what one would expect classically of a small charged sphere rotating about some axis, but the analogy can only be taken so far. It is, for example, not possible to make an electron “spin faster (or slower)” to alter the length of its projection onto an axis. This projection is always the same; it is a characteristic feature of the electron. What we are dealing with is an intrinsic property of the electron, like its mass $m_e$, that does not depend on its motion (or anything else); for this reason it is often referred to as the intrinsic angular momentum of the electron, but, unlike its classical counterpart, it is quantized, that is, can take only two discrete values.

Not only the electron, but every particle (elementary particle, atom, molecule, etc.) in quantum mechanics is supplied with some sort of intrinsic angular momentum. We will briefly describe the general situation (for more details see, for example, Chapter 11 of [Eis], or Chapters 14 and 17 of [Bohm]). The basic idea is that these particles exhibit behaviors that mimic what one would expect of angular momentum, but that cannot be accounted for by any orbital motion of the particle and, for a given particle, are always the same. To quantify these behaviors every particle is assigned a spin quantum number $s$. The allowed values of $s$ are

$0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \ldots, \frac{n-1}{2}, \ldots,$

where $n = 1, 2, 3, 4, 5, \ldots$. Intuitively, one might think of $n$ as the number of dots that appear on the photographic plate if a beam of such particles is sent through a Stern-Gerlach apparatus. According to this scheme an electron has spin $\frac{1}{2}$ ($n = 2$). For a particle with spin 0 there is just one dot, that is, there is no deflection at all. Particles with half-integer spin $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ are called fermions, while those with integer spin $0, 1, 2, \ldots$ are called bosons. Fermions and bosons have very different physical characteristics and play very different roles in particle physics (for more on this see Chapter 9 of [Nab5]). Among the elementary fermions, particles of spin $\frac{1}{2}$ are by far the principal players. Indeed, one must look long and hard to find an elementary particle of higher half-integer spin. The best know examples are the so-
called *baryons* which have spin $\frac{3}{2}$, but you dare not blink if you’re looking for one of these since their mean lifetime is about $5.63 \times 10^{-24}$ seconds. Among the bosons, the very recently observed *Higgs boson* has spin 0, whereas the conjectured, but not yet observed *graviton* has spin 2. The *photon* has spin 1, but it is massless and so does not quite fit into the $m > 0$ picture we have been discussing.

We have seen that the classical vector $\mathbf{S}$ used to describe the rotational angular momentum does not travel well into the quantum domain where it simply does not behave the way one expects a vector to behave. Nevertheless, it is still convenient to collect together the quantities $S_x, S_y$, and $S_z$, measured, for example, by a Stern-Gerlach apparatus aligned along the $x$-, $y$- and $z$-axes, and refer to the triple

$$\mathbf{S} = (S_x, S_y, S_z)$$

as the *spin vector*. Quantum theory decrees that, for a particle with spin quantum number $s$, the only allowed values for the “components” $S_x, S_y$, and $S_z$ are

$$-s\hbar, -(s-1)\hbar, \ldots, (s-1)\hbar, s\hbar. \quad (A.34)$$

In particular, for a spin $\frac{1}{2}$ particle such as the electron there are only two possible values so, for example,

$$S_z = \pm \frac{\hbar}{2}.$$

With this synopsis of the general situation behind us we will return to the particular case of spin $\frac{1}{2}$. We know that the classical picture of the electron as a tiny spinning ball cannot describe what is actually observed so we must look for another picture that can do this. Whatever this picture is it must be a quantum mechanical one so we are looking for a Hilbert space $\mathcal{H}$ and some self-adjoint operators on it to represent the observables $S_x, S_y$, and $S_z$. Previously we represented the state of the electron by a wave function $\psi(x, y, z)$ that is in $L^2(\mathbb{R}^3)$, but we now know that the state of a spin $\frac{1}{2}$ particle must depend on more that just $x, y$, and $z$ since this alone cannot tell us which of the two paths an electron is likely to follow in a Stern-Gerlach apparatus; we say “likely to” because we can no longer hope to know more than probabilities. What we would like to do is isolate some appropriate notion of the “spin state” of the particle that will provide us with the information we need to describe these probabilities. Now, we know that the only possible values of $S_z$ are $\pm \frac{\hbar}{2}$. By analogy with the classical situation one might view this as saying that the spin vector $\mathbf{S}$ can only be either “up” or “down”, but nothing in-between. This suggests that we consider wave functions

$$\psi(x, y, z, \sigma) \quad (A.35)$$

that depend on $x, y, z$, and an additional discrete variable $\sigma$ that can take only two values, say, $\sigma = 1$ and $\sigma = 2$ (or, if you prefer, $\sigma = \text{up}$ and $\sigma = \text{down}$). Then $|\psi(x, y, z, 1)|^2$ would represent the probability density for locating the electron at $(x, y, z)$ with $S_z = \frac{\hbar}{2}$ and similarly $|\psi(x, y, z, 2)|^2$ is the probability density for locat-
ing the electron at \((x, y, z)\) with \(S_z = -\frac{h}{2}\). Stated this way it sounds a little strange, but notice that this is precisely the same as describing the state of the electron with two functions \(\psi_1(x, y, z) = \psi(x, y, z, 1)\) and \(\psi_2(x, y, z) = \psi(x, y, z, 2)\) and this is what we will do. Specifically, we will identify the wave function of a spin \(\frac{1}{2}\) particle with a (column) vector

\[
\begin{pmatrix}
\psi_1(x, y, z) \\
\psi_2(x, y, z)
\end{pmatrix},
\]

where \(\psi_1\) and \(\psi_2\) are in \(L^2(\mathbb{R}^3)\) and

\[
\int_{\mathbb{R}^3} (|\psi_1(x, y, z)|^2 + |\psi_2(x, y, z)|^2) \, d\mu = 1
\]
because the probability of finding the electron somewhere with either \(S_z = \frac{h}{2}\) or \(S_z = -\frac{h}{2}\) is 1. The Hilbert space is therefore

\[
\mathcal{H} = L^2(\mathbb{R}^3; \mathbb{C}^2) \cong L^2(\mathbb{R}^3) \otimes \mathbb{C}^2.
\]

Now we must isolate self-adjoint operators on \(\mathcal{H}\) to represent the observables \(S_x, S_y\), and \(S_z\). Since these observables represent an intrinsic property of a spin \(\frac{1}{2}\) particle, independent of \(x, y\), and \(z\), we will want the operators to act only on the spin coordinates 1 and 2 and the action should be constant in \((x, y, z)\). Thus, we are simply looking for \(2 \times 2\) complex, self-adjoint (that is, Hermitian) matrices or, stated otherwise, self-adjoint operators on \(\mathbb{C}^2\). Since the only possible observed values are \(\pm \frac{h}{2}\), these must be the eigenvalues of each matrix. There are, of course, many such matrices floating around and we must choose three of them. Our choice is motivated by the desire to keep spin angular momentum and orbital angular momentum on the same formal footing since, classically at least, they really are the same thing.

We accomplish this by insisting that the operators satisfy the same commutation relations as those of the orbital angular momentum.

Recall that the generators \(M_{23} = J_1, M_{31} = J_2,\) and \(M_{12} = J_3\) in \(\mathfrak{p}\) can be realized as the components of the orbital angular momentum operators with \(\hbar = 1\) and that these satisfy the commutation relations (2.27), that is,

\[
[J_j, J_k] = i \epsilon_{jkl} J_l, \quad j, k = 1, 2, 3.
\]

Including a factor of \(\hbar\) in each operator these become

\[
[J_j, J_k] = i \hbar \epsilon_{jkl} J_l, \quad j, k = 1, 2, 3.
\]

Thus, we need to find operators \(S_1, S_2,\) and \(S_3\) on \(\mathbb{C}^2\) each of which has eigenvalues \(\pm \frac{h}{2}\) and that satisfy

\[
[S_j, S_k] = i \hbar \epsilon_{jkl} S_l, \quad j, k = 1, 2, 3.
\]

As it happens, this is quite easy. We let \(\sigma_j, \ j = 1, 2, 3,\) be the Pauli spin matrices (Exercise 1.2.6) and define
A.5 Spin

\[ S_j = \frac{\hbar}{2} \sigma_j, \quad j = 1, 2, 3. \quad (A.37) \]

**Exercise A.5.1.** Show that \( S_1, S_2, \) and \( S_3 \) given by \((A.37)\) satisfy the required conditions.

**Exercise A.5.2.** Define the operator \( S^2 \) on \( \mathbb{C}^2 \) by \( S^2 = S_1^2 + S_2^2 + S_3^2 \) and show that

\[
S^2 = \left( \frac{1}{2} \right) \left( \frac{1}{2} + 1 \right) \hbar^2 \text{id}_{\mathbb{C}^2}.
\]

**Note:** There actually is a point to writing \( \frac{3}{4} \) in this peculiar way as we shall soon see.

In order to make some direct contact with the material at the end of Section 2.8 we recommend the following exercise.

**Exercise A.5.3.** Let \( \mathcal{D}^{(1/2)} \) be the spin \( \frac{1}{2} \) representation of SU(2) on \( \mathbb{C}(\mathcal{D}^{(1/2)}) = \mathbb{C}^2 \) and let \( M_{23}, M_{31}, \) and \( M_{12} \) be the generators of rotations in \( p. \) Show that

\[
\begin{align*}
\frac{\text{i} \hbar}{\text{dt}} \big|_{t=0} (e^{-\text{i}utM_{23}}) &= S_1, \\
\frac{\text{i} \hbar}{\text{dt}} \big|_{t=0} (e^{-\text{i}utM_{31}}) &= S_2, \\
\frac{\text{i} \hbar}{\text{dt}} \big|_{t=0} (e^{-\text{i}utM_{12}}) &= S_3.
\end{align*}
\]

**Hint:** Remark 2.5.1 and Exercise 2.4.5.

The \( S_j = \frac{\hbar}{2} \sigma_j, j = 1, 2, 3, \) are operators on \( \mathbb{C}^2, \) but they give rise to operators \( \hat{S}_j = \text{id}_{L^2(\mathbb{R}^3)} \otimes S_j, j = 1, 2, 3, \) on the Hilbert space

\[
L^2(\mathbb{R}^3) \otimes \mathbb{C}^2 \cong L^2(\mathbb{R}^3) \otimes \mathcal{D}^{(1/2)}.
\]

Specifically,

\[
\begin{align*}
\hat{S}_1 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} &= \frac{\hbar}{2} \sigma_1 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} \psi_2(x,y,z) \\ -i\psi_1(x,y,z) \end{pmatrix}, \\
\hat{S}_2 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} &= \frac{\hbar}{2} \sigma_2 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} -i\psi_2(x,y,z) \\ \psi_1(x,y,z) \end{pmatrix}, \\
\hat{S}_3 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} &= \frac{\hbar}{2} \sigma_3 \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} \psi_1(x,y,z) \\ -\psi_2(x,y,z) \end{pmatrix}.
\end{align*}
\]
We call $\hat{S}_j$, $j = 1, 2, 3$, the spin operators for spin $\frac{1}{2}$ particles, although the terminology is often applied to the $S_j$, $j = 1, 2, 3$, themselves. The physical observables corresponding to these operators are the analogue of the classical spin angular momentum and are referred to either as the spin components or the intrinsic angular momentum components of the spin $\frac{1}{2}$ particle. These all have the same eigenvalues $\pm \frac{\hbar}{2}$ and the largest of these eigenvalues $\frac{\hbar}{2}$ is the spin of the particle (the term “spin $\frac{1}{2}$” tacitly assumes a factor of $\hbar$ or a choice of units in which $\hbar = 1$). In light of Exercise A.5.3 this coincides with the spin parameter for $D(1/2)$ introduced in Section 2.8 when the units are chosen so that $\hbar = 1$.

The scheme is precisely the same for particles of arbitrary spin $s = j/2$. These have wave functions with $2s + 1$ components corresponding to the $2s + 1$ “dots” on the Stern-Gerlach apparatus so that the Hilbert space is $\mathcal{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^{2s+1}$. The spin operators $S_1, S_2, S_3$ on $\mathbb{C}^{2s+1}$ are again required to satisfy the commutation relations (A.36) of angular momentum and have eigenvalues that are the observed values (A.34) of the spin components. One then notices that $\mathcal{C}(D(s)) \cong \mathbb{C}^{2s+1}$ and that a set of such matrices is given by

$$S_1 = i\hbar \left. \frac{d}{dt} [\mathcal{D}^{(s)}(e^{-i\Theta \hat{M}_{23}})] \right|_{t=0},$$

$$S_2 = i\hbar \left. \frac{d}{dt} [\mathcal{D}^{(s)}(e^{-i\Theta \hat{M}_{31}})] \right|_{t=0},$$

$$S_3 = i\hbar \left. \frac{d}{dt} [\mathcal{D}^{(s)}(e^{-i\Theta \hat{M}_{12}})] \right|_{t=0}.$$

We will have no need to write these out explicitly, but for those who would like to see $S_1, S_2$, and $S_3$ in one more case we recommend the following exercise for spin $s = 1$.

**Exercise A.5.4.** Define $3 \times 3$ matrices $S_1, S_2,$ and $S_3$ by

$$S_1 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_2 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

1. Show that $S_1, S_2,$ and $S_3$ satisfy the commutation relations (A.36).
2. Show that each $S_j$, $j = 1, 2, 3$, has eigenvalues $-\hbar, 0, \hbar$.
3. Show that

$$S^2 = S_1^2 + S_2^2 + S_3^2 = (1)(1+1)\hbar^2 \mathbb{C}^{3}.$$

4. Check at least one of the following.

$$S_1 = i\hbar \left. \frac{d}{dt} [\mathcal{D}^{(1)}(e^{-i\Theta \hat{M}_{23}})] \right|_{t=0},$$
A.5 Spin

\[ S_2 = i\hbar \frac{d}{dt} \left[ (D^{(1)}(e^{-iM_2 \Omega})) \right]_{t=0}, \]

\[ S_3 = i\hbar \frac{d}{dt} \left[ (D^{(1)}(e^{-iM_3 \Omega})) \right]_{t=0}. \]

We point out that, generalizing Exercise A.5.2 and Exercise A.5.4 (3), one finds that, for any \( s \),

\[ S^2 = S_1^2 + S_2^2 + S_3^2 = s(s+1)\hbar^2 \mathcal{C}^{2s+1} \]

The corresponding spin operators on the Hilbert space

\[ \mathcal{H} = L^2(\mathbb{R}^3) \otimes \mathcal{C}^{2s+1} \equiv L^2(\mathbb{R}^3) \otimes \mathcal{C}(\mathcal{D}^{(s)}) \]

are then

\[ \hat{S}_j = \text{id}_{L^2(\mathbb{R}^3)} \otimes \hat{S}_j, \quad j = 1, 2, 3. \]
References


Alex. Alexandrov, A.D., On Lorentz Transformations, Uspehi Mat. Nauk, 5, 3(37), 1950, 187 (in Russian)


References


GN. Gel’fand, I.M. and M.A. Naimark, On the Embedding of Normed Rings into the Ring of Operators in Hilbert Space, Mat. Sbornik 12, 1943, 197-213.


References

Iyen. Iyengar, K.S.K., A New Proof of Mehler’s Formula and Other Theorems on Hermitean Polynomials, Proc. Indian Acad. Sci., Section A, 10 (3), 1939, 211-216. (See http://repository.ias.ac.in/29465/)


References


Index

(\{v\}, [w]_{P(\mathbb{C})}, 26
Ad, 11
Ad_{\rho}, 11
Aut(G), 1
Aut(P(\mathbb{C})), 26
Aut(q), 11
H, 59
H_{\rho}, 40
H_{\alpha}, 10
Ind_{\rho}^{\mathbb{C}}(x), 32
L(x), 56
L_{\rho,\alpha}, 40
N \approx H, 33
P \times \mathbb{C} \ni, 30
R = (R_{\rho})_{\alpha, \beta = 0,1,2,3}, 53
S^1, 2
[ , ]_{L}, 5
P^2, 88
W^2, 89
\Lambda = (A^{\rho})_{\alpha, \beta = 0,1,2,3}, 47, 52
\Lambda_{\alpha}, 60
\Lambda(\theta), 54
L_{\nu}^{1}, 4, 47, 53
P_{\nu}, 47, 58
\Sigma(x^1, x^2, x^3), \tilde{\Sigma}(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3), \ldots, 44
A_{\nu}, 5
\mathcal{E}(\mathbb{D})/\mathbb{Z}^2, 23
\mathcal{E}_{\rho}(x_0), \mathcal{E}_{\rho}^{\nu}(x_0), 50
\mathcal{E}_{\nu}(x_0), \mathcal{E}_{\nu}^{\rho}(x_0), 50
\mathbb{D}/\mathbb{Z}^2, 23
\mathcal{E}(x_0), 51
\mathcal{E}(x_0), 51
\mathbb{M}, 43, 48
\mathcal{O}, \mathcal{O}, \ldots, 44
\mathcal{O}_{0}, 40
\mathcal{O}_{\alpha}, 10
S(x^0, x^1, x^2, x^3), \tilde{S}(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3), \ldots, 46
\Sigma, \Sigma^*, 50
\mathcal{U}(\mathbb{C}), 14
\mathcal{U}(q), 86
\mathcal{U}^{-1}(\mathbb{C}), 27
\langle , \rangle, 49
C^*, 2
P(\mathbb{C}), 24
R^*, 1
R^{1,3}, 51
\eta = (\eta_{\rho})_{\alpha, \beta = 0,1,2,3}, 47, 49
\eta^{-1} = (\eta_{\rho}^{\nu})_{\alpha, \beta = 0,1,2,3}, 49
\mathbb{N}, 38
\kappa: SL(2, \mathbb{C}) \rightarrow \mathcal{L} \mathcal{L}, 61, 63
\frak{iso}(3), 36
\nu, 72
\mathbb{A}, 3
\pi_{\nu}, 24
\sigma_{\nu}, 10
\sigma_{j, j = 1, 2, 3, 18}
\tau_{\nu}, 55
\{e_0, e_1, e_2, e_3\}, 48
Ad, 11
ad_{\nu}, 11
1-parameter group of diffeomorphisms, 130
1-particle space, 115
4-accleration, 57
4-velocity, 57
abstract Schrödinger equation, 156
action functional, 124
adjoint action, 12
adjoint representation of \mathcal{U}(\mathbb{C}), 87
adjoint representation of \mathcal{U}(q), 12
admissible basis, 51
admissible observer, 44
algebra of classical observables, 135, 140
angular momentum
classical, 131
intrinsic, 168
orbital, 164
rotational, 164
total, 164
angular momentum operator
orbital, 120
spin, 121
total, 121
angular momentum operators
orbital, 163
anticommutation relations, 19
associated Legendre functions, 80
associative algebra, 8
associative algebra
commutator bracket, 9
commutator Lie algebra structure, 9
ideal, 8
ideal generated by $S$, 9
subalgebra, 8
unit, 8
unital, 8
automorphism group of $P(H)$, 26, 155
automorphism group of a Lie group, 1
automorphism of $P(H)$, 26, 155
automorphism of Lie groups, 1
Bargmann’s Theorem, 105
base space, 29
boost, 54
Born-von Neumann formula, 151
boson, 168
bracket, 4
bundle space, 29
canonical 1-form, 133
canonical commutation relations
for classical mechanics, 141
for quantum mechanics, 159
canonical coordinates, 133, 140
canonical map, 142
canonical symplectic form, 133
cartan’s magic formula, 136
Casimir invariants, viii
Casimir invariants of $p$, 90
causal automorphism, 46
causality assumption, 46
Change of Variables Formula, 100
character, 38
character group, 38
of $R^3$, 39
classical canonical commutation relations, 141
classical harmonic oscillator Hamiltonian, 135
classical observable, 135, 140
Closed Subgroup Theorem, 2
commutation relations, 19
$iso(3)$, 37
commutator bracket, 9
commutator Lie algebra structure, 9
configuration space, 123, 133
conjugate momentum, 125
conjugation representation, 18
conservation law
classical mechanics, 126
conservation of energy, 137, 142
conserved quantity, 125, 137
and the Poisson bracket, 142
angular momentum, 131
energy, 142
contragredient action, 93, 103
covector, 133
covering group, 13
covering map, 13, 63
covering space, 13
critical curve, 125
critical point, 125
Darboux Theorem
finite dimensional, 139
linear, 139
derivation, 11
dispersion, 151
double cover
universal, 63
dual group, 38
duration of a timelike vector, 55
Einstein energy-momentum relation, 95
Einstein summation convention, ix
elsewhere, 51
energy, 153
equivalent projective representations, 28
Euler-Lagrange equations, 125, 132
and Newton’s Second Law, 128
event, 43, 49
evolution operator, 153
expectation value, 151
expected value, 151
exponential map, 9
fermion, 168
fiber, 29
Fizeau procedure, 45
flow of a vector field, 130
frame of reference, 46
free group action, 10
<table>
<thead>
<tr>
<th>Index</th>
<th>187</th>
</tr>
</thead>
<tbody>
<tr>
<td>free particle, 128</td>
<td></td>
</tr>
<tr>
<td>general linear group, 2</td>
<td></td>
</tr>
<tr>
<td>generators of a Lie algebra, 36</td>
<td></td>
</tr>
<tr>
<td>generators of boosts, 66</td>
<td></td>
</tr>
<tr>
<td>generators of rotations, 37</td>
<td></td>
</tr>
<tr>
<td>generators of translations, 37</td>
<td></td>
</tr>
<tr>
<td>GL(n, C), 2</td>
<td></td>
</tr>
<tr>
<td>GL(n, R), 2</td>
<td></td>
</tr>
<tr>
<td>global section</td>
<td></td>
</tr>
<tr>
<td>of a principal H-bundle, 30</td>
<td></td>
</tr>
<tr>
<td>graviton, 169</td>
<td></td>
</tr>
<tr>
<td>group action, 10</td>
<td></td>
</tr>
<tr>
<td>group representation</td>
<td></td>
</tr>
<tr>
<td>induced, 29</td>
<td></td>
</tr>
<tr>
<td>semi-direct products, 33</td>
<td></td>
</tr>
<tr>
<td>Haar measure, 1</td>
<td></td>
</tr>
<tr>
<td>Hamilton’s equations, 134, 136, 142</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian, 133, 140</td>
<td></td>
</tr>
<tr>
<td>classical harmonic oscillator, 135</td>
<td></td>
</tr>
<tr>
<td>on $L^2(\mathbb{R}^n)$, 156</td>
<td></td>
</tr>
<tr>
<td>quantum, 153</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian flow, 134</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian Noether Theorem, 144</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian system, 140</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian vector field, 133, 136, 140, 142</td>
<td></td>
</tr>
<tr>
<td>harmonic oscillator</td>
<td></td>
</tr>
<tr>
<td>classical Hamiltonian, 135</td>
<td></td>
</tr>
<tr>
<td>Heisenberg equation, 158</td>
<td></td>
</tr>
<tr>
<td>Heisenberg picture, 158</td>
<td></td>
</tr>
<tr>
<td>Heisenberg Uncertainty Principle, 152</td>
<td></td>
</tr>
<tr>
<td>Hermitian inner product, 3</td>
<td></td>
</tr>
<tr>
<td>Higgs boson, 169</td>
<td></td>
</tr>
<tr>
<td>Hilbert bundle, 30, 31</td>
<td></td>
</tr>
<tr>
<td>section, 31</td>
<td></td>
</tr>
<tr>
<td>homogeneous manifold, 12</td>
<td></td>
</tr>
<tr>
<td>homogeneous space, 12</td>
<td></td>
</tr>
<tr>
<td>ideal, 8</td>
<td></td>
</tr>
<tr>
<td>ideal generated by $S$, 9</td>
<td></td>
</tr>
<tr>
<td>indefinite inner product, 4</td>
<td></td>
</tr>
<tr>
<td>induced representation, 29, 32, 42</td>
<td></td>
</tr>
<tr>
<td>infinitesimal symmetry, 126, 143</td>
<td></td>
</tr>
<tr>
<td>inhomogeneous rotation group, 35</td>
<td></td>
</tr>
<tr>
<td>inhomogeneous SL(2, C), 64, 103</td>
<td></td>
</tr>
<tr>
<td>internal semi-direct product, 34</td>
<td></td>
</tr>
<tr>
<td>intertwine, 15</td>
<td></td>
</tr>
<tr>
<td>intrinsic angular momentum, 168, 172</td>
<td></td>
</tr>
<tr>
<td>invariant under rotation, 129</td>
<td></td>
</tr>
<tr>
<td>irreducible representation, 15</td>
<td></td>
</tr>
<tr>
<td>ISO(3), 35</td>
<td></td>
</tr>
<tr>
<td>isolated quantum system, 153</td>
<td></td>
</tr>
<tr>
<td>isomorphic Lie algebras, 5</td>
<td></td>
</tr>
<tr>
<td>isomorphism of Lie groups, 1</td>
<td></td>
</tr>
<tr>
<td>isomorphism of projectivizations, 26</td>
<td></td>
</tr>
<tr>
<td>isotropy subgroup, 10</td>
<td></td>
</tr>
<tr>
<td>Jacobi Identity, 136, 141</td>
<td></td>
</tr>
<tr>
<td>Jacobi identity, 4</td>
<td></td>
</tr>
<tr>
<td>kinetic energy operator, 160</td>
<td></td>
</tr>
<tr>
<td>Lagrangian, 124</td>
<td></td>
</tr>
<tr>
<td>symmetry of, 126</td>
<td></td>
</tr>
<tr>
<td>left action, 10</td>
<td></td>
</tr>
<tr>
<td>left invariant vector field, 5</td>
<td></td>
</tr>
<tr>
<td>Leibniz rule, 11</td>
<td></td>
</tr>
<tr>
<td>Leibniz rule, 11</td>
<td></td>
</tr>
<tr>
<td>Lie algebra, 4</td>
<td></td>
</tr>
<tr>
<td>complex, 4</td>
<td></td>
</tr>
<tr>
<td>complexification, 5</td>
<td></td>
</tr>
<tr>
<td>exponential map on, 9</td>
<td></td>
</tr>
<tr>
<td>isomorphism, 5</td>
<td></td>
</tr>
<tr>
<td>of a Lie group, 5</td>
<td></td>
</tr>
<tr>
<td>of GL(n, C), 7</td>
<td></td>
</tr>
<tr>
<td>of GL(n, R), 7</td>
<td></td>
</tr>
<tr>
<td>of O(n), 7</td>
<td></td>
</tr>
<tr>
<td>of O(p, q), 8</td>
<td></td>
</tr>
<tr>
<td>of SO(n), 7</td>
<td></td>
</tr>
<tr>
<td>of SO(p, q), 8</td>
<td></td>
</tr>
<tr>
<td>of SU(n), 8</td>
<td></td>
</tr>
<tr>
<td>of the Lorentz group, 65</td>
<td></td>
</tr>
<tr>
<td>of the U(n), 8</td>
<td></td>
</tr>
<tr>
<td>Lie algebra automorphism, 5</td>
<td></td>
</tr>
<tr>
<td>Lie algebra cohomology, 28</td>
<td></td>
</tr>
<tr>
<td>Lie algebra homomorphism, 5</td>
<td></td>
</tr>
<tr>
<td>Lie algebra isomorphism, 5</td>
<td></td>
</tr>
<tr>
<td>Lie algebra of operators, 82</td>
<td></td>
</tr>
<tr>
<td>Lie algebra representation, 5</td>
<td></td>
</tr>
<tr>
<td>Lie group, 1</td>
<td></td>
</tr>
<tr>
<td>closed subgroups of, 2</td>
<td></td>
</tr>
<tr>
<td>continuous homomorphisms of, 2</td>
<td></td>
</tr>
<tr>
<td>isomorphism, 1</td>
<td></td>
</tr>
<tr>
<td>representation, 1</td>
<td></td>
</tr>
<tr>
<td>Lie groups, 1</td>
<td></td>
</tr>
<tr>
<td>Lie subalgebra, 5</td>
<td></td>
</tr>
<tr>
<td>lift of a projective representation, 28</td>
<td></td>
</tr>
<tr>
<td>light travel time, 45</td>
<td></td>
</tr>
<tr>
<td>Linear Darboux Theorem, 139</td>
<td></td>
</tr>
<tr>
<td>little group, 40</td>
<td></td>
</tr>
<tr>
<td>local section</td>
<td></td>
</tr>
<tr>
<td>of a principal H-bundle, 30</td>
<td></td>
</tr>
<tr>
<td>locally isomorphic Lie groups, 7</td>
<td></td>
</tr>
<tr>
<td>Lorentz group</td>
<td></td>
</tr>
<tr>
<td>general, 4</td>
<td></td>
</tr>
<tr>
<td>proper, 4</td>
<td></td>
</tr>
<tr>
<td>proper, orthochronous, 4</td>
<td></td>
</tr>
</tbody>
</table>
Lorentz inner product, 49
Lorentz transformation
  general, 52
  orthochronous, 53
  proper, 52
  special, 55

Mackey machine, vii, 37, 41
  and ISL(2, C), 107
Mackey’s Theorem, 41
magnetic dipole, 164
magnetic moment
  orbital, 164
  rotational, 165
  total, 165
mass hyperboloid, 93
mass operator, 118
matrix Lie group, 3
matrix mechanics, 148
measure
  pushforward, 102
measurement, 147, 153
Minkowski inner product, 49
Minkowski spacetime, 48
moment map, 144
momentum, 128
  conjugate, 125
momentum map, 144
momentum operator, 162
momentum space, viii, 92
natural coordinates on TM, 123
Noether’s Theorem, 127, 144
null cone, 50
  future, 50
  past, 50
null curve, 56
  future directed, 56
  past directed, 56
null vector, 49

O(n), 3
O(p, q), 3
observable
  classical, 135
  quantum, 149
orbit, 10
orbit space, 30
orbital angular momentum, 164
orbital magnetic moment, 164
orthogonal group, 3
orthogonal transformation of M, 52
orthogonal vectors in M, 50
orthonormal basis for M, 49
past directed, 50
path space, 124
Pauli spin matrices, 18
Pauli-Lubanski vector, 88
phase space
  in classical mechanics, 133
  in Hamiltonian mechanics, 137
photon, 169
Poincaré algebra, 72
Poincaré group, 58
Poincaré-Birkhoff-Witt Theorem, 86
Poisson algebra, 136
Poisson bracket, 136, 141
Poisson commutes, 137
position operator, 161
positive energy, 94
Postulates of Quantum Mechanics
  QM1, 148
  QM2, 149
  QM3, 150
  QM4, 153
potential, 156
  spherically symmetric, 128
potential energy operator, 161
principal H-bundle
  trivial, 29
principal bundle, 29, 30
principal H-bundle, 29
Principle of Least Action, 125
Principle of Stationary Action, 125
probability amplitude, 149
projection, 29
  of a principal bundle, 29
projective representation, vii, 24, 28
  lift, 28
projectivization of Ω, 149
projectivization of a Hilbert space, 24
proper time function, 57
proper time length, 56
proper time separation, 55
pushforward measure, 102
quantization, 161
quantum bracket, 159
quantum canonical commutation relations, 159
quantum field, vii
quantum symmetry group, 155
quantum system, 147
  state space, 149
  symmetry of, 26, 154
quasi-invariant measure, 32
ray in a Hilbert space, 24
realization of a Lie algebra, 84
reducible representation, 15
regular semi-direct product, 40
relativistic invariance, viii
relativistic angular momentum, 77
relativistic scalar field, 116
Relativity Principle, 48
representation
and left action, 14
carriers, 14
dimension, 14
induced, 32, 42
invariant subspace, 14, 15
irreducible, 15
of a group, 14
reducible, 15
spinor, 23
trivial, 21
Wigner, 116
rest frame, 115
Reversed Schwartz Inequality, 55
Reversed Triangle Inequality, 56
right action, 10
Robertson Uncertainty Relation, 152
Rodrigues’ formula, 80
rotation group, 3, 129
exponential map on, 129
Lie algebra of, 129
rotation subgroup of \( \mathbb{C}^+ \), 54
rotational angular momentum, 164
rotational magnetic moment, 165
scalar field
relativistic, 116
scalar field on \( \mathbb{R}^n \), 116
Schrödinger equation, 148
abstract version, 156
physicist’s version, 156
Schrödinger picture, 158
Schur’s Lemma, 15
for associative algebras over \( \mathbb{C} \), 91
infinite-dimensional version, 15
section
global, 30
local, 30
of a Hilbert bundle, 31
semi-direct product, 33, 35
internal, 34
regular, 40
representations, 33
semi-direct product of Lie groups, 37
semi-orthogonal group, 3
signature of Minkowski spacetime, ix
\( \text{SL}(2, \mathbb{C}) \)
conjugation representation, 20
standard representation, 20
\( \text{SL}(n, \mathbb{C}) \), 3
\( \text{SL}(n, \mathbb{R}) \), 3
\( \text{SO}(3) \)
geometrical description, 10
\( \text{SO}(n) \), 3
\( \text{SO}(p, q) \), 4
space of an admissible observer, 51
spacelike curve, 56
spacelike vector, 49
special linear group, 3
special Lorentz transformation, 55
hyperbolic form, 55
special orthogonal group, 3
special semi-orthogonal group, 4
special unitary group, 3
special unitary group \( \text{SU}(2) \), 17
spherical harmonics, 80
spherically symmetric potential, 128
spin, 172
of an \( \text{SU}(2) \) representation, 24
spin \( \frac{1}{2} \), 168
spin angular momentum operator, 121
spin components, 172
spin model of \( \mathbb{R}^3 \), 19
spin operators, 172
spin quantum number, 168
spin vector, 169
spinor
contravariant, 23
rank \( m \), 23
\( \text{SL}(2, \mathbb{C}) \), 24
\( \text{SU}(2) \), 23
spinor field on \( \mathbb{R}^n \), 116
spinor map, 63
spinor representation
\( \text{SU}(2) \), 23
state
classical mechanical system, 123, 133
in quantum mechanics, 148
state space, 123
in quantum mechanics, 149
stationary curve, 125
stationary point, 125
Stern-Gerlach experiment, 165
structure constants, 7
structure group, 29
\( \text{SU}(2) \)
defining representation, 17
representation
spin, 24
spinor representation, 23
standard representation, 17
\( \text{SU}(n) \), 3
SU(2), 17
irreducible representations, 17
subalgebra, 8
surface measure, 100
symmetry, 126
quantum system, 26, 154
symmetry group, 143
of a Lagrangian, 127
physical interpretation, 84
quantum, 155
symplectic diffeomorphism, 142
symplectic form, 137
canonical, 133
symplectic gradient, 136, 140
symplectic manifold, 137
compact, 138
is even dimensional, 138
is orientable, 138
symplectic vector field, 142
symplectomorphism, 142
time axis, 51
time cone, 50
future, 50
past, 50
time orientation, 50, 51
timelike curve, 56
future directed, 56
past directed, 56
timelike vector, 49
timelike worldline, 56
total angular momentum, 164
total energy
classical mechanics, 133
in quantum mechanics, 153
symplectic geometry, 140
total magnetic moment, 165
total orbital angular momentum, 78
total relativistic energy, 94
total space, 29
transition amplitude, 153
transition probability, 26, 153
transitive group action, 10
trivial principal $H$-bundle, 29
trivial representation, 14, 21	
twin paradox, 56
U(n), 3
uncertainty relations, 152
Robertson, 152
unit ray, 25, 149
unitarily equivalent representations, 15
unitary group, 3
unitary group representation, 14
unitary representation, 14
strongly continuous, 14
weakly continuous, 14
universal covering group, 63
universal double cover, 63
universal enveloping algebra, viii, 85
basis, 87
variance, 151
wave function, 148
wave mechanics, 148
weight of spinor representation, 23
Wigner representation, 116
Wigner representation of mass $m$ and spin $j/2$, 116
Wigner rotation, 67
Wigner symmetry, 26, 154
Wigner’s Theorem on Symmetries, 26, 154
worldline, 43
worldline of a material particle, 56