1 Introduction

Throughout the history of physics a key tool in the development of physical laws has been the infinitesimal. Infinitesimals entered the scene with the development of the calculus by Newton and Leibniz and went on to be used, to mention some of the most notable cases, in Euler’s derivation of the laws of hydrodynamics, in the development of analytical mechanics by Euler, Lagrange and Hamilton and in Riemann’s differential geometry. Infinitesimals are a key conceptual tool for constructing laws in which local behaviour determines global behaviour. This is a natural way of deriving physical laws, which explains why infinitesimal arguments are so popular in the history of physics. A limitation of these methods, however, was that they were not rigorous, a fact that was more of a concern to 19th Century mathematicians, rather than the physicists. The attempts to rigorize the calculus, which culminated in the development of set theory, replaced infinitesimals with limits and rendered infinitesimals mathematical fictions, since, in a classical context, physical models use archimedean fields so the only number that has the properties that infinitesimals have is 0. This did not stop physicists from using infinitesimal arguments in the development of physical laws, however, since such methods much more natural then the use of limits.

In order to use infinitesimals in physical arguments, a way of using infinitesimals even though they were not available in the model was required. There are two main ways of using infinitesimals in a classical context. The usual approach was to consider an infinitesimal to merely be a small but finite quantity—small enough that powers of some degree are negligible but large enough that it is non-zero and, hence, exists in the model. Essentially, in such arguments certain quantities are considered to be nilsquare infinitesimals as a first order approximation (or $(k+1)$-square as a $k$-th order approximation). Consequently, such quantities are only approximately infinitesimals leading to not entirely rigorous arguments. Another way of bringing infinitesimals into arguments in a classical context is to formally add them to the model. Formal infinitesimals $\epsilon$ are introduced by adjuction of elements to a ring or a field $R$:

$$R(\epsilon) = R[x]/(x^k), \quad \epsilon^k = 0,$$

where $R[x]$ is the polynomial ring with coefficients in $R$ and $(x^k)$ is the ideal generated by $x^k$ for some finite $k$. In this case, the arguments are rigorous, but not true to the intuitions behind infinitesimal arguments since the infinitesimals are artificially added and there are only finitely many of them. In both cases, the infinitesimals considered
are nilpotent infinitesimals, i.e. infinitesimals in which some power is zero, which are the natural infinitesimals for physical argumentation.

Recent developments in logic and category theory, however, have breathed life back into the infinitesimal. Both nonstandard analysis and smooth infinitesimal analysis contain genuine infinitesimals and do so in a consistent and fully rigorous way.\footnote{Nonstandard analysis shifts from archimedean fields to non-archimedean fields and smooth infinitesimal analysis switches from fields to rings, in a manner in which we consider in the next section.} This is interesting from the point of view of arguments in physics, since these systems provide a way of making infinitesimal arguments both rigorous and true to the intuitions behind the original use of infinitesimal arguments. That the infinitesimals of smooth infinitesimal analysis are nilpotent, makes them particularly interesting from this point of view, because of the prevalence of nilpotent infinitesimals in physical argumentation. In this paper I look at the development of the framework for dynamics in smooth infinitesimal analysis. I begin by developing the basic mathematical machinery and then use this to discuss the classical groups in the context of SIA. I then discuss the basic framework of classical Hamiltonian dynamics, which may be applied to a variety of physical theories formulated in a smooth setting, including unitary evolution in quantum mechanics.

2 Smooth Infinitesimal Analysis

Before we can enter into an examination of dynamics, we will first need to develop some of the basic theory of smooth infinitesimal analysis (SIA). A model of SIA is a topos $\mathcal{S}$ in which all maps are smooth maps. Any particular $\mathcal{S}$ is called a smooth world. We assume throughout that we are working in a sufficiently rich intuitionistic set theory so that the basic set theoretic constructions are possible.

2.1 Basic Principles

The fundamental object in $\mathcal{S}$ is an ordered unitary commutative ring $\mathbb{R}$, which we will call the Euler real line. $\mathbb{R}$ is required to be an algebra over the rational numbers. $\mathbb{R}$ has as a sub-object a commutative semigroup (under multiplication) $\Delta = \{ \epsilon \in \mathbb{R} | \epsilon^2 = 0 \}$, called the (nil-square) infinitesimals of $\mathbb{R}$. $\mathbb{R}$ is required to be a ring and not a field so that $\Delta$ does not reduce to $\{0\}$. The order relation $\leq$ on $\mathbb{R}$ is reflexive and transitive, compatible with the unitary ring structure\footnote{To say that the order relation $\leq$ is compatible with the unitary ring structure means that the following statements are true: $\forall x, y, z \in \mathbb{R} \ (x \leq y) \rightarrow (x + z \leq y + z)$, $\forall x, y \in \mathbb{R} \ (0 \leq x \land 0 \leq y) \rightarrow (0 \leq x \cdot y)$, $0 \leq 1 \land \neg(1 \leq 0)$.} and has the property

$$\forall \epsilon \in \Delta \ 0 \leq \epsilon \land \epsilon \leq 0. \tag{2.1}$$

From the object $\Delta$, also known as the first-order infinitesimals, we may form collections of infinitesimals of arbitrary orders. If $\epsilon_1, \epsilon_2 \in \Delta$, then $(\epsilon_1 + \epsilon_2)^2 = 2\epsilon_1\epsilon_2$, which is not
necessarily zero, but we do have that \((\epsilon_1 + \epsilon_2)^3 = 0\), so that \(\eta = \epsilon_1 + \epsilon_2\) is a nilcube infinitesimal. Thus, we may define

\[
\Delta_2 = \{\epsilon_1 + \epsilon_2|\epsilon_1, \epsilon_2 \in \Delta\},
\]
as the collection of second-order (or nilcube) infinitesimals. This can be iterated arbitrarily so that

\[
\Delta_k = \left\{ \sum_{i=1}^{k+1} \epsilon_i \mid \epsilon_j \in \Delta \right\},
\]
is the collection of \(k\)-th order infinitesimals. The collection formed from all the \(\Delta_k\) for arbitrary \(k\)—the additive closure of \(\Delta\)—is called the collection of nilpotent infinitesimals.

A closed interval \([a, b]\) of \(\mathbb{R}\) is defined as

\[
[a, b] \equiv \{x \in \mathbb{R} \mid a \leq x \land x \leq b\}.
\]

If we define \(<\) to be the relation defined by \(x < y \leftrightarrow x \leq y \land x \neq y\), then we may define an open interval \((a, b)\) of \(\mathbb{R}\) as \((a, b) \equiv \{x \in \mathbb{R} \mid a < x \land x < b\}\). A sub-object \(J\) of \(\mathbb{R}\) is said to be microstable if

\[
\forall \epsilon \in \Delta \quad a \in J \rightarrow a + \epsilon \in J.
\]

**Proposition 2.1** Every closed interval \([a, b]\) is microstable.

**Proof:** Let \(x \in [a, b]\), so that \(a \leq x \land x \leq b\), and let \(\epsilon \in \Delta\) be arbitrary but fixed. Since \(0 \leq \epsilon\) by (2.1), it follows that \(x \leq x + \epsilon\), so that \(a \leq x + \epsilon\) by transitivity. Similarly, \(\epsilon \leq 0\) entails that \(x + \epsilon \leq b\). Therefore, \(x + \epsilon \in [a, b]\). □

The first basic principle is the first axiom of SIA:

**Axiom 1 (Principle of Microaffineness (Kock-Lawvere Axiom))** For any map \(g: \Delta \rightarrow \mathbb{R}\), there exists a unique \(b \in \mathbb{R}\) such that,

\[
\forall \epsilon \in \Delta \quad g(\epsilon) = g(0) + b\epsilon.
\]

The principle of microaffineness (PMA) has a simple consequence, which is the second basic principle of SIA:

**Theorem 2.1 (Principle of Microcancellation)** For any \(a, b \in \mathbb{R}\), if \(a\epsilon = b\epsilon\) for all \(\epsilon \in \Delta\), then \(a = b\).

**Proof:** (Bell [2]) Suppose that \(a\epsilon = b\epsilon\) for all \(\epsilon \in \Delta\). Consider the function \(g: \Delta \rightarrow \mathbb{R}\) such that \(g(\epsilon) = a\epsilon\). Then we have that \(g(\epsilon) = a\epsilon = b\epsilon\), but the uniqueness of the slope given by the principle of microaffineness entails that \(a = b\). □

### 2.2 Vectors, R-modules and C-modules

The closest equivalent of a real vector space in \(\mathbb{S}\) is an R-module. We will refer to the elements \(v\) of an R-module as vectors. An R-module \(E\) is defined to be Euclidean if for every function \(f: \Delta \rightarrow E\) there is a unique vector \(v \in E\) such that

\[
\forall \epsilon \in \Delta \quad f(\epsilon) = f(0) + v\epsilon.
\]

Since \(\mathbb{S}\) is Cartesian closed, it contains all finite products, which entails that \(\mathbb{R}^n\) and \(\text{End}(\mathbb{R}^n) \cong \mathbb{R}^{n \times n}\) are in \(\mathbb{S}\). \(\mathbb{R}^n\) and \(\mathbb{R}^{n \times n}\) have a natural R-module structure and can be thought to be generated by a basis. We will assume, unless stated otherwise, some arbitrary basis is given for the vectors of an R-module.
A particular collection of vectors is worthy of mention here. For any given \( R \)-module, the collection of infinitesimal vectors with mutually cancelling infinitesimals as components is defined to be
\[
\Delta(n) = \{(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) | \epsilon_i\epsilon_j = 0\}.
\]
\( \Delta(n) \) is the collection of \textit{microvectors} of dimension \( n \). Geometrically, the microvectors are the infinitesimal vectors that point in a definite direction in an \( n \)-dimensional \( R \)-module. Equivalently, we may think of a microvector \( \mathbf{v} \) as an \( n \)-tuple
\[
\mathbf{v} = (x_1, x_2, \ldots, x_n)\epsilon,
\]
where \( x_1, \ldots, x_n \in R \) and \( \epsilon \in \Delta \).

We now prove that the \( R \)-modules of interest have the desired property of being Euclidean.

**Proposition 2.2** \( \mathbb{R}^n \) is a Euclidean \( R \)-module.

**Proof:** Let \( f : \Delta \to \mathbb{R}^n \) and \( \epsilon \in \Delta \) be arbitrary. Since \( f(\epsilon) = (x_1, \ldots, x_n) \), \( f \) induces a collection of maps \( f_i : \Delta \to R, \epsilon \mapsto x_i \), called the \textit{components} of \( f \). By PMA, for each \( f_i \) we have \( f_i(\epsilon) = f_i(0) + b_i\epsilon \), where each \( b_i \) is unique given \( \epsilon \). It then follows that \( f(\epsilon) = f(0) + b\epsilon \), where \( b = (b_1, \ldots, b_n) \) is unique. \( \square \)

**Theorem 2.2** Let \( V \) be an \( R \)-module and let \( E \) be Euclidean. Then the \( R \)-module \( L(V, E) \) is Euclidean.

**Proof:** Let \( f, g \in L(V, E) \). If we let \( f + g \) be defined by \((f + g)(v) = f(v) + g(v)\) and let \( cf \) be defined by \((cf)(v) = c(f(v)) \) for \( v \in V \) and \( c \in R \), then \( L(V, E) \) becomes an \( R \)-module. Let \( h : \Delta \to L(V, E) \) and \( \epsilon \in \Delta \) be arbitrary. Then for any \( v \in V, h(\epsilon)(v) \in E \). Now, letting \( h_v(\epsilon) = h(\epsilon)(v) \), we define a function from \( \Delta \) to \( E \). Since \( E \) is Euclidean, \( h_v(\epsilon) = h_v(0) + b_v\epsilon \), where \( b_v \) is unique. Thus, \( h(\epsilon)(v) = h(0)(v) + b_v\epsilon \). Letting \( v \) vary defines a function \( b : V \to E \), which is linear since, for any \( v_1, v_2 \in V \) and \( c_1, c_2 \in R \)
\[
h(\epsilon)(c_1v_1 + c_2v_2) = c_1h(\epsilon)(v_1) + c_2h(\epsilon)(v_2) = h(0)(c_1v_1 + c_2v_2) + (c_1b_{v_1} + c_2b_{v_2})\epsilon =
\]
\[
h(0)(c_1v_1 + c_2v_2) + b_{v_1+c_2v_2}\epsilon.
\]
Moreover, since each \( b_v \) is unique given \( \epsilon \) and \( v \), given \( \epsilon \) the function obtained by varying \( v \) is unique. Therefore \( b(\epsilon) = h(0) + be \), with \( b \) unique given \( \epsilon \). \( \square \)

**Corollary 2.1** \( \mathbb{R}^{n \times n} \) is Euclidean.

**Proof:** The result follows immediately from the fact that \( L(\mathbb{R}^n, \mathbb{R}^n) = \text{End}(\mathbb{R}^n) \) and \( \text{End}(\mathbb{R}^n) \cong \mathbb{R}^{n \times n} \). \( \square \)

An \( R \)-module of particular interest is the \( R \)-module which is isomorphic to \( \mathbb{R}^2 \) but has \( \{1, i\} \) as its standard basis, where \( i = \sqrt{-1} \). This \( R \)-module is denoted by \( \mathbb{C} \) an is called the \textit{complex line} or the \textit{complex numbers}. A particular element \( z \) of \( \mathbb{C} \), or complex number, is written as \( z = x + iy \), where \( x, y \in R \) and \( x \) and \( y \) are called, respectively, the real and imaginary parts of \( z \). If we define the product of two complex numbers \( z_1 = x_1 + iy_1 \) and \( z_2 = x_2 + iy_2 \) to be \( z_1z_2 = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1) \), then \( \mathbb{C} \) becomes a unitary commutative ring, with identity 1 and the addition operation inherited from the \( R \)-module structure. This enables us to consider \( \mathbb{C} \)-modules as well, which are the SIA equivalent of complex vector spaces. Since \( \mathbb{S} \) is Cartesian closed, \( \mathbb{C}^n \) and \( \text{End}(\mathbb{C}^n) \cong \mathbb{C}^{n \times n} \) are in \( \mathbb{S} \). Infinitesimal complex numbers, \( \xi = \epsilon + i\eta, \epsilon\eta = 0 \), also called \textit{microcomplex numbers}, are a special case of a microvector. The collection of microcomplex numbers is denoted \( \Delta' \). We may then define the collections of complex \( n \)-dimensional microvectors \( \Delta'(n) \) composed of \( n \)-tuples of mutually cancelling microcomplex numbers.
A $\mathbb{C}$-module $E$ is defined to be **Euclidean** if for any $f: \Delta \to E$ there is a unique vector $v \in E$ such that $f(\epsilon) = f(0) + be$. This enables us to prove the following two propositions, the proofs of which are *mutatis mutandis* the same as their real equivalents:

**Proposition 2.3** $\mathbb{C}^n$ is a Euclidean $\mathbb{C}$-module.

**Theorem 2.3** Let $V$ be an $\mathbb{C}$-module and let $E$ be a Euclidean $\mathbb{C}$-module. Then the $\mathbb{R}$-module $L(V, E)$ is Euclidean.

**Corollary 2.2** $\mathbb{C}^{n \times n}$ is Euclidean.

2.3 Differentiation

2.3.1 Functions of a Single Variable

The definition of the derivative in SIA is a straightforward consequence of the principle of microaffineness. Let $f: J \to \mathbb{R}$ be an arbitrary function defined on a microstable sub-object $J$ of $\mathbb{R}$. For fixed $x \in J$, $f(x + \epsilon)$ is a function from $\Delta$ to $\mathbb{R}$. By PMA, there is a unique $b_x$ depending on $x$ such that

$$f(x + \epsilon) = f(x) + b_x \epsilon \quad (2.2)$$

for all $\epsilon \in \Delta$. The variation of $x$ induces a function $f': J \to \mathbb{R}$, $x \mapsto b_x$. The function $f'$ is defined to be the derivative of $f$. The form of equation (2.2) that we will use often is

$$f(x + \epsilon) = f(x) + f'(x)\epsilon, \quad (2.3)$$

where $x \in J$ and $\epsilon \in \Delta$ are arbitrary. Equation (2.3) is called the fundamental equation of the differential calculus (FEDC) in $\mathbb{S}$. $f'(x)$ is also called the slope at $x$ of the curve determined by $f$, which is motivated by the fact that $f(x + \epsilon) - f(x) = f'(x)\epsilon$ can be thought of as an infinitesimal tangent to the curve defined by $f$, which coincides with $f$ at $x$. We may now state and prove the following proposition:

**Proposition 2.4** Let $f: J \to \mathbb{R}$, $g: J \to \mathbb{R}$ and $c \in \mathbb{R}$ be arbitrary. Then $(cf + g)' = cf' + g'$ and the Leibniz law $(fg)' = f'g + fg'$ holds.

**Proof:** Let $\epsilon \in \Delta$ by arbitrary. Then we have that $(cf + g)(x + \epsilon) = cf(x + \epsilon) + g(x + \epsilon) = (cf + g)(x) + (cf' + g')(x)\epsilon$, but we also have $(cf + g)(x + \epsilon) = (cf + g)(x) + (cf + g)'(x)\epsilon$. Thus, $(cf + g)' = cf' + g'$ by PMC. Now, under the same conditions we have that $(fg)(x + \epsilon) = f(x + \epsilon)g(x + \epsilon) = (f(x) + f'(x)\epsilon)(g(x) + g'(x)\epsilon) = (fg)(x) + ((f'g)(x) + (fg')(x))\epsilon)$, but we also have that $(fg)(x + \epsilon) = (fg)(x) + (fg)'(x)\epsilon$. The Leibniz law then follows by PMC. \( \square \)

Since the function $f$ in the definition of the derivative is arbitrary, every function $f: J \to \mathbb{R}$ in $\mathbb{S}$ is differentiable. This entails that the process of taking derivatives can be iterated indefinitely, so that by defining the $n$th derivative of $f$ recursively by

$$f^{(n-1)}(x + \epsilon) = f^{(n-1)}(x) + f^{(n)}(x),$$

entails that every real-valued function on the reals in $\mathbb{S}$ is smooth. Since closed intervals are microstable, *a fortiori* all functions defined on closed intervals are smooth.

---

\(^3\)Since $J$ is microstable, if $x$ is in $J$, so is $x + \epsilon$.  

Following the above line of reasoning, we are immediately able to define a derivative for functions from a microstable sub-object $J$ of $R$ to a Euclidean $R$-module or $C$-module $E$. Let $f: R \to E$, then for fixed $x \in J$ we have that $f(x + \epsilon) = f(x) + b(x)\epsilon$ by the definition of a Euclidean $R$- or $C$-module. Variation of $x$ then induces a function $f': J \to E$, $f'(x) = b(x)$, which is defined to be the derivative of $f$. Since the derivative is defined for an arbitrary function $f: J \to E$, all such functions are smooth. A path is defined to be a function $f$ from a microstable sub-object $J$ of $R$ to a $R$- or $C$-module $V$. In some cases the image of the function in $V$ will be referred to as the path. Thus, we immediately have the following:

**Proposition 2.5** All paths in Euclidean $R$- and $C$-modules are smooth.

In particular, all paths in $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{R}^{n \times n}$ and $\mathbb{C}^{n \times n}$ are smooth.

### 2.3.2 Functions of Several Variables

For the remainder of this subsection, $V$ will be an arbitrary $R$-module and $E$ will be an arbitrary Euclidean $R$-module and $f: V \to E$ an arbitrary function. Let $x, v \in V$ be arbitrary but fixed. Then $f(x + ve)$ defines a function from $\Delta$ to $E$. Since $E$ is Euclidean, there exists a unique $b \in E$ such that

$$f(x + ve) = f(x) + be.$$

The unique $b$ is denoted $(\partial_v f)(x)$ and is called the derivative of $f$ in the direction $v$.

**Theorem 2.4** For every $x \in V$, $\partial_v f(x)$ is a linear function of $v$.

**Proof:** See Lavendhomme [8].

The linear function $df(x): V \to E, v \mapsto df(x)(v) = \partial_v f(x)$ defined by the directional derivative, is called the differential of $f$ with respect to $x$. The differential of $f$ can be thought of as an object that measures the change in $f$ in an infinitesimal neighborhood of the point $x \in V$. By feeding $df$ a vector $v \in V$, $df(x)(v)e$ then gives the infinitesimal change in $f$ in the direction $v$.

**Proposition 2.6** Let $H$ be a Euclidean $R$-module, let $f: V \to E, g: V \to E$ and $h: E \to F$ be arbitrary functions, and let $df$, $dg$ and $dh$ be their differentials. Let $c \in R$. Then the following are true:

1. $d(cf + g) = cdf + dg$;
2. $d(h \circ f)(x) = dh(f(x)) \circ df(x)$.

**Proof:** We omit the proof.

Let $V = \mathbb{R}^n$ and let $e_i$ be a unit vector along the $x^i$ coordinate axis. Then, the partial derivative with respect to $x^i$ is defined as

$$\frac{\partial f}{\partial x^i}(x) \equiv \partial_{x^i} f(x).$$

Let $v \in \mathbb{R}^n$, so that $v = v^i e_i$. Then, by the linearity of the directional derivative, we have that

$$\partial_v f(x) = \frac{\partial f}{\partial x^i}(x)v^i. \quad (2.4)$$

---

4Here, and throughout the paper, we use the summation convention on repeated indices, unless otherwise specified.
If we let $\pi^i: \mathbb{R}^n \to \mathbb{R}$ be the $i$-th coordinate (projection) function, then we have that
\[
\pi^i(x + v\epsilon) = x^i + v^i\epsilon,
\]
(2.5)
but we also have that
\[
\begin{align*}
\pi^i(x + v\epsilon) &= \pi^i(x) + \partial_v \pi^i(x)\epsilon \\
&= x^i + d\pi^i(v)\epsilon \\
&= x^i + dx^i(v)\epsilon,
\end{align*}
\]
(2.6)
(2.7)
(2.8)

Letting $dx^i \equiv d\pi^i(x)$. Then by PMC we have that
\[
dx^i(v) = v^i.
\]
(2.9)

Using equation (2.4) we may obtain
\[
df(x) = \frac{\partial f}{\partial x^i}(x)dx^i.
\]
(2.10)

Equation (2.10) enables us to regard the $dx^i$ as a basis for differentials.

The above multivariable theory can be developed mutatis mutandis for $\mathbb{C}$-modules.

2.4 Integration

The integration theory in smooth infinitesimal analysis requires an additional axiom:

**Axiom 2 (Integration Principle)** For any $f: [0, 1] \to \mathbb{R}$ there is a unique $g: [0, 1] \to \mathbb{R}$ such that $g' = f$ and $g(0) = 0$.

From this axiom we may prove the theorem that generalizes to arbitrary intervals (for a proof see Bell [2]).

**Theorem 2.5** For any $f: [a, b] \to \mathbb{R}$ there is a unique $g: [a, b] \to \mathbb{R}$ such that $g' = f$ and $g(a) = 0$.

For all $x \in [a, b]$, the value of the unique function $g$ such that $g' = f$ and $g(a) = 0$ is denoted by
\[
\int_a^x f(t)dt \quad \text{or} \quad \int_a^x f.
\]

We then have the following proposition (which we state without proof):

**Proposition 2.7** For all functions $f: [a, b] \to \mathbb{R}$ and $g: [a, b] \to \mathbb{R}$ and $c \in \mathbb{R}$:

1. $\int_a^b (f + g) = \int_a^b f + \int_a^b g$;
2. $\int_a^b c \cdot f = c \cdot \int_a^b f$;
3. $\int_a^b f' = f(b) - f(a)$;
4. $\int_a^b f' \cdot g = [f(x) \cdot g(x)]_a^b - \int_a^b f \cdot g'$.

To treat differential equations involving complex valued real functions, we will need the following extension of the integration principle.
Proposition 2.8 \textit{(Quasi-complex Integration Principle (QCIP))} For any \( f: [a, b] \to \mathbb{C} \) there is a unique \( g: [a, b] \to \mathbb{C} \) such that \( g' = f \) and \( g(0) = 0 \).

Proof: Let \( f: [a, b] \to \mathbb{C} \) be a function. Since \( f(t) = x(t) + iy(t) \) and \( x(t) \) and \( y(t) \) are both functions from \( [a, b] \) to \( \mathbb{R} \), there exists unique functions \( g_1: [a, b] \to \mathbb{C} \) and \( g_2: [a, b] \to \mathbb{C} \) such that \( g'_1(t) = x(t) \), \( g_1(0) = 0 \) and \( g'_2(t) = y(t) \), \( g_2(0) = 0 \) by theorem 2.5. If we let \( g(t) = g_1(t) + ig_2(t) \), then \( g(0) = 0 \) and \( g'(t) = f(t) \). Since \( g \) is unique because \( g_1 \) and \( g_2 \) are unique, this proves the proposition. \( \square \)

3 Lie Groups, and Their Lie Algebras and Representations

In classical mathematics, the description of the dynamical evolution of physical systems is often given as a path in some smooth geometric space, real or abstract. Since this evolution is smooth and reversible, the possible evolution operations naturally have a Lie group structure, which makes the theory of Lie groups a useful theoretical tool in dynamics. The adaptation of this theory to SIA is a highly non-trivial matter.\(^5\) Thus, rather than discussing general Lie groups, we will work with a restricted notion of Lie group.

To motivate our notion of Lie group, let us consider classical Lie groups. Classical Lie groups are groups that are also smooth manifolds such that the differential structure is compatible with the group structure, \( i.e. \) the group operations are smooth. Since we want to avoid the difficulties concerned with considering an SIA analogue of general Lie groups, we want a restricted definition of a Lie group that serves our present purposes. Since all maps in SIA are smooth maps, we need not worry about the compatibility of the ‘manifold structure’ and the group structure, since it will be guaranteed that the group operations are smooth. Thus, we really are interested in groups with a differential structure. Thus, I suggest that an appropriate definition of a Lie group in this context is the following:

A Lie group is a Euclidean \( \mathbb{R} \)- or \( \mathbb{C} \)-module that is also a group.

With this, the Euler spaces \( \mathbb{R}^n \) and \( \mathbb{C}^n \) are Lie groups under addition. The other important groups that this definition includes are the SIA versions of the classical linear groups, \( i.e. \) the subgroups of \( GL_n(\mathbb{R}) \) and \( GL_n(\mathbb{C}) \), which are multiplicative groups. We will refer to the matrix Lie groups as the \textit{linear Lie groups}. From this point on, to streamline this section we will formally consider only complex linear Lie groups, which proceeds almost identically to the real case.\(^6\)

With our notion of Lie group in hand we may lay down some of the basic definitions for Lie groups. A \textit{Lie subgroup} \( H \) of a Lie group \( G \) is a subobject of \( G \) that is also a group and a Euclidean \( \mathbb{R} \)- or \( \mathbb{C} \)-module. Let \( G \) and \( H \) be Lie groups. A map \( f: G \to H \) is a homomorphism of Lie groups if it is a homomorphism of groups. A homomorphism is an \textit{isomorphism} if there exists a map \( g: H \to G \) such that \( f \circ g = 1_H \), \( g \circ f = 1_G \). A \textit{representation} of a Lie group \( G \) on a \( \mathbb{R} \)- or \( \mathbb{C} \)-module \( M \) is a Lie group homomorphism \( \rho: G \to GL(M) = Aut(M) \).

We notice that a Lie group homomorphism is naturally an action.

\(^5\) Lavendhomme [8] defineds a Lie group as a microlinear group, which requires a complex network of definitions to flesh out.

\(^6\) The proof of theorem 3.2 differs slightly in the real case because there is an additional possibility to consider for the roots of equation (3.35), but this does not significantly change the course of the proof.
We now turn to our consideration of the linear Lie groups. If $L(C^n, C^n)$ is the associative $C$-algebra of linear maps from $C^n$ to itself, then Aut($C$) is defined to be the group of invertible elements of $L(C^n, C^n)$. Similarly, Aut($R$) is defined to be the group of invertible elements of $L(R^n, R^n)$. We will exploit the fact that $L(C^n, C^n) \cong C^{n \times n}$ and work directly with the general linear group of invertible matrices, denoted $GL_n(C)$. The classical matrix groups can be specified as the locus $S$ of zeros of a finite set of polynomial functions,

$$S = \{x \in C^{n \times n} \mid f_i(x) = 0\},$$

i.e. as a complex (or real) algebraic variety. This is because the matrix groups are defined by placing restrictions on the determinant, and the determinant is a polynomial function. We will call the algebraic varieties $S$ real (or complex) matrix varieties.

Our focus here will be on paths in linear Lie groups and paths in Euler spaces generated by paths in linear Lie groups. A path in a linear Lie group is a path $g(t)$ lying entirely in $S$. Such a path will also be a path in the matrix space $C^{n \times n}$ and since $C^{n \times n}$ is Euclidean, we have that

$$g(t + \epsilon) = g(t) + g'(t)\epsilon,$$

where $g'(t)$ is called (a) tangent to $S$ at $x$. Let $x = g(t)$ and $v = g'(t)$. Now, since $g(t)$ is a path in $S$, we must have

$$f_i(g(t + \epsilon)) = f_i(x + \epsilon v) = f_i(x) + \frac{\partial f_i}{\partial x_k} v_k \epsilon.$$  \quad (3.2)

Since $g(t)$ lies entirely in $S$, $f_i(x) = 0$ and $f_i(g(t + \epsilon)) = 0$ so that

$$\frac{\partial f_i}{\partial x_k} v_k = 0$$  \quad (3.3)

for paths in $S$ by PMC. If, rather, we wish to infinitesimally generate a path in $S$ given a point $x \in S$, we want to know which vectors $v$ keep will keep the path in $S$, i.e. the vectors $v$ for which $f_i(x + \epsilon v) = 0$ for all $\epsilon \in \Delta$. The vectors $v$ that have this property are called the infinitesimal tangents to $S$ at $x$. That the infinitesimal tangents do indeed generate the paths in $S$ follows from the following proposition:

**Proposition 3.1** Let $x \in S$, where $S$ is a matrix variety. A vector $v$ is tangent to $S$ at $x$ iff $v$ is an infinitesimal tangent to $S$ at $x$.

**Proof:** For the forward direction, suppose that $v$ is tangent to $S$ at $x$, then there is a path $h(t)$ in $S$ such that $h(t + \epsilon) = x + \epsilon v$. But since $h$ is in $S$, $f_i(h(t + \epsilon)) = f_i(x + \epsilon v) = 0$. For the reverse direction, suppose that $f_i(x + \epsilon v) = 0$. Let $h: [0, 0] \to R^{n \times n}$ be the path such that $h(0) = x$ and $h'(t) = v$. Then $f_i(h(0 + \epsilon)) = f_i(x + \epsilon v) = 0$ so that $h$ lies in $S$. Therefore, $v$ is tangent to $S$ at $x$. \square

The collection $T_x(G)$ of all tangent vectors at a point $x$ in a linear Lie group $G$ is the tangent space of $G$ at $x$.

**Theorem 3.1** Let $x$ be a point in a complex linear Lie group $G$. Then $T_x(G)$ is a $C$-module.
Proof: We will regard the group \(G\) as a complex algebraic variety. Since \(x \in S\), \(f_i(x + 0\epsilon) = f_i(x) = 0\), so that \(0 \in T_x(G)\) by proposition 3.1. Let \(c \in \mathbb{C}\) and \(v, w \in T_x(G)\). Since each \(f_i\) is smooth, we have that
\[
f_i(x + (cv + w)\epsilon) = f_i(x) + \frac{\partial f_i}{\partial x_k} (cv + w)_k \epsilon
= 0 + c \frac{\partial f_i}{\partial x_k} v_k \epsilon + \frac{\partial f_i}{\partial x_k} w_k \epsilon.
\] (3.4)
But by equation (3.3) we have that
\[
\frac{\partial f_i}{\partial x_k} v_k = 0 = \frac{\partial f_i}{\partial x_k} w_k,
\] (3.5)
so that \(f_i(x + (cv + w)\epsilon) = 0\). Thus, \(cv + w \in T_x(G)\) by proposition 3.1. Letting \(c = -1\) and \(w = 0\) shows that if \(v \in T_x(G)\), so is \(-v\). It follows that \(T_x(G)\) is an \(\mathbb{C}\)-module. □

A tangent space of particular importance is the tangent space of the identity \(T_1(G)\), which is called the Lie algebra of the linear Lie group \(G\). The Lie algebra is, in a concrete sense, the possible slopes of curves in \(G\) at the identity. \(T_1(G)\) becomes a \(\mathbb{C}\)-algebra by adding to the \(\mathbb{C}\)-module structure a product operation
\[
[v, w] = vw - wv,
\] (3.7)
called the Lie Bracket.

**Proposition 3.2** Let \(v, w \in T_1(G)\). Then \([v, w] \in T_1(G)\).

Proof: Since \(v, w \in T_1(G)\), for any \(\epsilon, \eta \in \Delta, (1 + v\epsilon), (1 - v\epsilon), (1 + w\eta), (1 - w\eta) \in G\). Since each of these are in \(G\), their product must also be in \(G\), so we have that
\[
(1 + v\epsilon)(1 + w\eta)(1 - v\epsilon)(1 - w\eta) = 1 + (vw - wv)\epsilon\eta
\] (3.8)
is in \(G\). Since \(\epsilon\) and \(\eta\) are arbitrary and the map \((\epsilon, \eta) \mapsto \epsilon\eta\) is onto, it follows that \([v, w] = vw - wv \in T_1(G)\).

The Lie bracket is the infinitesimal form of the group commutator \(ghg^{-1}h^{-1}\). It follows immediately from the \(\mathbb{C}\)-module structure of \(T_1(G)\) that the Lie bracket is a skew-symmetric bilinear form, i.e. a bilinear form with the property that \([v, w] = -[w, v]\). The Lie bracket is not associative, but it does satisfy the Jacobi identity:
\[
\forall u, w, v \in T_1(G) \quad [u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0,
\] (3.9)
which is easily verified by direct calculation. Any \(\mathbb{C}\)-module that has a product operation that is bilinear, skew-symmetric and satisfies the Jacobi identity is called a Lie \(\mathbb{C}\)-algebra. Thus, \(T_1(G)\) is a Lie \(\mathbb{C}\)-algebra in the formal sense. The Lie \(\mathbb{C}\)-algebra of a linear Lie group \(G\) will be denoted by \(\mathfrak{g}\).

We now turn to the consideration of paths in linear Lie groups. In traditional treatments paths are considered, as a heuristic device, to be infinitesimally generated, i.e. they are generated as integral curves of vector fields. In smooth infinitesimal analysis, however, paths actually are infinitesimally generated as the infinitesimal generators are
in the group and any infinitesimal change along the path is accomplished by an infinitesimal motion. In fact, we can derive the differential equation for paths using infinitesimal arguments. To see this, consider a path \( g : [a, b] \to G \). Since \( g \) is also path in \( \mathbb{C}^{n \times n} \), which is Euclidean, we have

\[
g(t + \epsilon) = g(t) + g'(t)\epsilon.
\]

Thus, an infinitesimal shift along the path is accomplished by an infinitesimal motion in the direction \( g'(t) \). There is also another way to accomplish an infinitesimal shift along the path, by moving the image \( p \) of a given point \( t \) in \( [a, b] \) by an infinitesimal amount. Since the identity element \( 1 \) of the group leaves the point fixed, an infinitesimal motion of the point will differ only infinitesimally from the identity. Thus, we need to multiply \( p \) by a group element of the form

\[
1 + \xi(t)\epsilon,
\]

where \( \xi(t) \in \mathfrak{g} \) is called the velocity of the path. We are assured that \( (1 + \xi(t)\epsilon) \in G \) because \( \xi(t) \) is required to be in \( \mathfrak{g} \). Applying (3.11) to \( g(t) \), we obtain

\[
g(t + \epsilon) = (1 + \xi(t)\epsilon)g(t). \tag{3.12}
\]

From equations (3.10) and (3.11) it follows that \( g'(t)\epsilon = \xi(t)g(t)\epsilon \) and, since \( \epsilon \) is arbitrary, that

\[
g'(t) = \xi(t)g(t) \tag{3.13}
\]

by PMC. Equation (3.13) is the differential equation of a path in a Lie group. Although equation (3.13) holds generally for classical Lie groups, the particular argument used here is only valid for linear Lie groups \( G \).

A special case of particular interest is where \( J = \mathbb{R} \) and \( g \) is a homomorphism such that \( t + s \mapsto g(s)g(t) \). Such paths are called one-parameter subgroups. Since \( g \) is a path in \( G \) and \( a \text{ fortiori} \) a path in \( \mathbb{C}^{n \times n} \), which is Euclidean, \( g(0) = 1 \) entails that

\[
g(\epsilon) = 1 + \xi\epsilon \tag{3.14}
\]

for some \( \xi \in \mathfrak{g} \), and

\[
g(t + \epsilon) = g(t) + g'(t)\epsilon. \tag{3.15}
\]

Now, \( g(t + \epsilon) = g(\epsilon)g(t) \), so that

\[
g(t + \epsilon) = (1 + \xi\epsilon)g(t) \tag{3.16}
\]

by (3.14). It follows, then, that \( g'(t)\epsilon = \xi g(t)\epsilon \) by (3.15). Thus, we obtain

\[
g'(t) = \xi g(t) \tag{3.17}
\]

by the principle of microcancellation (PMC). Equation (3.17), a special case of (3.13) for one-parameter subgroups, shows that all one-parameter subgroups have constant velocity. Stated otherwise, the velocity is an invariant under motion along the path.

We end this section with the following theorem.

**Theorem 3.2** There exists a unique solution to the initial value problem

\[
g'(t) = \xi g(t), \quad g(0) = 1. \tag{3.18}
\]
Proof: In order prove the the existence and uniqueness of solutions to equation (3.18) we must decouple the \( n^2 \) first-order linear differential equations. To simplify the notation, we will let \( X = g(t) \) and \( A = \xi \), so that (using the notation \( \dot{X} = X' \))

\[
\dot{X} = AX, \quad X(0) = 1. \tag{3.19}
\]

We make the observation that \( \frac{dX}{dt} = \frac{dX_i}{dt} = A_k^i X_j \), so that the \( j \)-th column of \( \frac{dX}{dt} \) depends only on the \( j \)-th column of \( AX \). Thus, equation (3.19) is equivalent to \( n \) equations of the form

\[
\dot{x}_i = Ax_i, \quad [x_i(0)]^j = \delta_i^j \tag{3.20}
\]

where \( x_i \in \mathbb{C}^n \) is a column vector. Now, every \( A \in \mathbb{C}^{n \times n} \) can be be expressed in the form

\[
A = QRQ^{-1}, \tag{3.21}
\]

where \( R \) is the rational canonical form for \( A \) and \( Q \) is an invertible matrix. In general, \( R \) is of the form

\[
R = \begin{pmatrix}
C_1 & O & \cdots & O \\
O & C_2 & \cdots & O \\
& & \ddots & \vdots \\
O & O & \cdots & C_k
\end{pmatrix}; \tag{3.22}
\]

where each \( O \) is a zero matrix of the appropriate size, and each \( C_j \) is a \( l \) dimensional matrix block of the form

\[
C_j = \begin{pmatrix}
0 & 0 & \cdots & 0 & -a_0 \\
1 & 0 & \cdots & 0 & -a_1 \\
0 & 1 & \cdots & 0 & -a_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & -a_{l-1}
\end{pmatrix}. \tag{3.23}
\]

We may now note the following two things. Firstly, if we let \( \bar{x}_i = Q^{-1}x_i \), then if \( \bar{x}_i \) is a solution of

\[
\dot{\bar{x}}_i = R\bar{x}_i, \quad [\bar{x}_i(0)]^j = [Q^{-1}]_j^i, \tag{3.24}
\]

then \( x_i = Q\bar{x}_i \) is a solution of equation (3.20). Thus, we may restrict our attention to the case where \( A \) is a rational canonical form. Secondly, we observe that if \( A \) is a rational canonical form, equation (3.20) splits into \( k \) independent equations of the form

\[
\dot{x}_{ik} = C_k x_{ik}, \quad [x_{ik}(0)]^j = c_j, \tag{3.25}
\]

for some \( b_j \). Thus, the general problem of solutions to equation (3.19) reduces to solutions of equation (3.25).

We will demonstrate existence and uniqueness of solutions to equation (3.25) by explicit construction. Suppose that \( C_k \) is a matrix of dimension \( l \), then we seek an \( l \) dimensional column vector \( x_{ik} \) as the solution. To clarify notation, we will let \( x \equiv x_{ik} \). Since \( C_k \) is a matrix of the form (3.23), we obtain from equation (3.25) the \( l \) equations

\[
\dot{x}^1 = -a_0 x^j \tag{3.26}
\]

\[
\dot{x}^j = x^{j-1} - a_{j-1} x^j, \quad j > 1, \tag{3.27}
\]

\footnote{Naturally, the repeated indices are not summed over here.}
with the initial conditions \( x^j(0) = c^j \). Letting \( j = l \) in equation (3.27) we obtain

\[
\dot{x}^l = x^{l-1} - a_{l-1}x^l. \tag{3.28}
\]

If we differentiate both sides of equation (3.28) then we obtain

\[
\ddot{x}^l = x^{l-2} - a_{l-2}x^l - a_{l-1}\dot{x}^l, \tag{3.29}
\]

\[
x^l = x^{l-1} - a_{l-1}x^l \tag{3.30}
\]

by equation (3.27). By iterating this process we obtain

\[
(x^l)^{(k)} = x^{l-k} - a_{l-k}x^l - \cdots - a_{l-1}(x^l)^{(k-1)}, \tag{3.31}
\]

which after \( l - 1 \) iterations yields the following homogeneous linear \( l \)-th order differential equation with constant coefficients:

\[
(x^l)^{(l)} + a_{l-1}(x^l)^{(l-1)} + \cdots + a_1x^l + a_0x^l = 0, \tag{3.32}
\]

where \( (x)^{(k)} = \frac{d^k x}{dt^k} \). The initial value of \( x^l \) is fixed to be \( x^l(0) = c^l \), and using equation (3.31), the initial conditions \( x^l(0) \) determine the initial values of the next \( l - 2 \) derivatives of \( x^l \) are determined recursively by

\[
(x^l)^{(l-k)}(0) = x^{l-k}(0) - a_{l-k}x^l(0) - \cdots - a_{l-1}(x^l)^{(k-1)}(0) \tag{3.33}
\]

\[
= c^{l-k} - a_{l-k}x^l(0) - \cdots - a_{l-1}(x^l)^{(k-1)}(0). \tag{3.34}
\]

To generate a solution we consider the trial solution \( x^l = e^{rt} \), \( r \in \mathbb{C} \). Via a substitution of this trial solution into equation (3.32), we obtain the following equation for \( r \):

\[
r^l + a_{l-1}r^{l-1} + \cdots + a_1r + a_0 = 0. \tag{3.35}
\]

There are two distinct possibilities. The first possibility is that this equation will have \( l \) distinct roots, so that the general solution to the differential equation (3.32) will be

\[
x^l = \sum_{i=1}^{l} b_i e^{r^i t}. \tag{3.36}
\]

If there are only \( m < l \) distinct roots, then if a given root \( r_p \) is repeated \( n \) times, then we still have the following \( n \) distinct solutions

\[
e^{r_p t}, \quad t e^{r_p t}, \quad \ldots \quad , t^{n-1} e^{r_p t}. \tag{3.37}
\]

In this case we obtain a modified version of equation (3.36). It does not affect the argument if we assume from here on that \( x^l \) has the form of equation (3.36) so we will do so from now on. We can differentiate equation (3.36) to obtain expressions for the derivatives of \( x^l \) up to \( (x^l)^{(k)} \), which with the initial values of the derivatives of \( x^l \) yield \( l \) equations in \( l \) unknowns which can be solved to uniquely determine the values of the \( b_i \). This yields a unique solution to the differential equation (3.32). From equation (3.27) we have that

\[
x^j = \dot{x}^{j+1} + a_j x^l, \quad j < l, \tag{3.38}
\]

which, using equation (3.36) and its derivatives, enables us to recursively determine the values of each of the \( x^j \). Since the derivative of a function is unique, each of the \( x^j \) is unique. Thus, we have obtained a unique solution to equation (3.25). Thus, we have obtained a unique solution to equation (3.19). \( \square \)
4 Smooth Infinitesimal Evolution

4.1 Classical Mechanics: Hamiltonian Mechanics and Canonical Transformations

Traditional treatments of analytical mechanics begin with a Lagrangian treatment and then develop the more theoretical Hamiltonian framework. Following this approach, let us consider a system with \( n \) degrees of freedom. The mechanics of a system is then treated by determining \( n \) generalized coordinates \( q_i \) and the Lagrangian \( L(q_i, \dot{q}_i, t) = T - V \) for the system, where \( T \) is the kinetic energy and \( V \) is the potential energy. The equation of motion is then determined by solving the Euler-Lagrange equations

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0,
\]

(4.1)

where \( \dot{q}_i \) is the first time derivative of \( q_i \). A generalized momentum, called the \textit{canonical momentum} or \textit{conjugate momentum}, is then defined to be

\[
p_i = \frac{\partial L}{\partial \dot{q}_i}.
\]

(4.2)

The Euler-Lagrange equations are derived from a least action principle (Hamilton's principle), which uses variational methods. The treatment of variational arguments in SIA is beyond the scope of the present work, so we will just assume given sets of equations without motivation.

The Hamiltonian framework can be obtained in one of two ways: from the Lagrangian framework via a Legendre transformation or independently from a least action principle. Since we are not considering variational methods, we will only review the former. The Hamiltonian \( H \) is generated by a Legendre transformation

\[
H(q, p, t) = \dot{q}_i p_i - L(q_i, \dot{q}_i, t).
\]

(4.3)

Taking the differential \( dH \) of \( H \) we obtain (by applying equation (2.10))

\[
dH = \dot{q}_i dp_i - \dot{p}_i dq_i - \frac{\partial L}{\partial t} dt,
\]

(4.4)

but since

\[
dH = \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial t} dt,
\]

(4.5)

we obtain the 2\( n + 1 \) relations:

\[
\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad -\dot{p}_i = \frac{\partial H}{\partial q_i},
\]

(4.6)

\[
-\frac{\partial L}{\partial t} = \frac{\partial H}{\partial t}.
\]

(4.7)

Equations (4.6) are called \textit{Hamilton’s equations}.

There are several things worthy of note concerning Hamilton’s equations. In contrast to the Euler-Lagrange equations, which are \( n \) second-order differential equations in \( n \) unknowns, Hamilton’s equations are \( 2n \) first-order equations in \( 2n \) unknowns. This highlights the fact that the coordinates and conjugate momenta are treated as independent variables in solving Hamilton’s equations. Not only are they independent, but the
equations (4.6) are nearly symmetric under interchange of \( q_i \) and \( p_i \). This leads to a much more abstract view where which of \( p_i \) and \( q_i \) is considered to be the coordinate and which the momentum becomes arbitrary. On this point of view it is just the form of the equations that is important and not the interpretation of the variables. Exploiting the near symmetry leads to a powerful notation called the *symplectic notation*. We define a phase space coordinate \( q \) (a 2\( n \)-dimensional column vector) such that

\[
[q]^i = q_i, \quad [q]^{i+n} = p_i,
\]

and a 2\( n \)-dimensional column vector \( \frac{\partial H}{\partial q} \) such that

\[
\begin{bmatrix}
\frac{\partial H}{\partial q}
\end{bmatrix}^i = \frac{\partial H}{\partial q_i}, \quad \begin{bmatrix}
\frac{\partial H}{\partial q}
\end{bmatrix}^{i+n} = \frac{\partial H}{\partial p_i}.
\]

Then, if we define the matrix

\[
J = \begin{bmatrix}
O & I \\
-I & O
\end{bmatrix},
\]

where \( I \) is the \( n \times n \) identity matrix and \( O \) is the \( n \times n \) zero matrix, we may write Hamilton’s equations in the compact (symplectic) form

\[
\dot{q} = J \frac{\partial H}{\partial q}.
\]

Since it is the form of Hamilton’s equations that is important, transformations of coordinates that preserve the form of Hamilton’s equations are of considerable theoretical interest. To treat this problem we may consider point transformations of \( p \) and \( q \), where \( q = (q_1, \ldots, q_n)^t \) and \( p = (p_1, \ldots, p_n)^t \),

\[
q' = q'_i(q, p, t) \quad (4.12)
\]
\[
p' = p'_i(q, p, t) \quad (4.13)
\]

for which there exists a function \( K(q', p', t) \) such that

\[
\dot{q}'_i = \frac{\partial K}{\partial p'_i}, \quad \dot{p}'_i = \frac{\partial K}{\partial q'_i}.
\]

The function \( K \) plays the role of the Hamiltonian in the new coordinate system. This is accomplished if we have

\[
p_i \dot{q}'_i - H = p'_i \dot{q}'_i - K + \frac{dF}{dt},
\]

where \( F \) is any transformation of the phase space coordinates. Equations (4.12) and (4.13) allow \( F \) to be expressed partly in terms of the old and partly in terms of the new coordinates. The function \( F \) becomes useful for expressing the precise form of the transformation only when it is expressed half in terms of the old and half in terms of the new coordinates, which enables it to act as a bridge between the two coordinates. In this capacity \( F \) is called the generating function of the transformation.

\[\text{This equation may be derived by requiring that both sets of coordinates satisfy the same variational principle (modified Hamilton’s principle), which gives a more general equation, viz. where the left hand side is multiplied by a scale factor } \lambda, \text{ but canonical transformations consider only } \lambda = 1. \text{ This is so since any transformation can be accomplished via a canonical transformation composed with a scale transformation, making the canonical transformations of primary concern.}\]
We may also treat the problem of canonical transformations using the symplectic notation. Let us restrict our attention to the case where the transformation does not depend explicitly on time and, hence, the Hamiltonian does not change. The new coordinates \( q'_i \) and \( p'_i \) define a column vector \( \mathbf{q}' \) such that
\[
\mathbf{q}' = \mathbf{q}(\mathbf{q}). \tag{4.16}
\]
From the chain rule we have that
\[
\dot{\mathbf{q}}' = M \dot{\mathbf{q}}, \quad [M]_{ij} = \frac{\partial q'_i}{\partial q_j}. \tag{4.17}
\]
By Hamilton’s equations (4.11), then, we have
\[
\mathbf{q}' = MJ \frac{\partial H}{\partial \mathbf{q}}. \tag{4.18}
\]
To obtain an expression purely in terms of \( Q' \), we may utilize the fact that
\[
\frac{\partial H}{\partial q} = \tilde{M} \frac{\partial H}{\partial q'}, \tag{4.19}
\]
which then yields
\[
\mathbf{q}' = MJ \tilde{M} \frac{\partial H}{\partial \mathbf{q}}. \tag{4.20}
\]
To ensure that the transformation is canonical, \( i.e. \) preserves Hamilton’s equations, we restrict attention to those transformations for which
\[
MJ \tilde{M} = J. \tag{4.21}
\]
Equation (4.21) is called the symplectic condition for a canonical transformation. The symplectic condition holds generally for all canonical transformations, which we will shortly show. Any matrix \( M \) is called a symplectic matrix. The collection of matrices that satisfy the symplectic condition form the real symplectic group \( Sp_n(R) \).

**Theorem 4.1** The symplectic condition is a necessary and sufficient condition for a transformation to be canonical.

We will prove this in two stages. That the symplectic condition is sufficient is basically trivial. Suppose that \( M \) is a transformation matrix satisfying the symplectic condition. From Hamilton’s equations and equation (4.19) we have that
\[
\dot{\mathbf{q}} = J \tilde{M} \frac{\partial H}{\partial \mathbf{q}'}, \tag{4.22}
\]
so that
\[
M \dot{\mathbf{q}} = MJ \tilde{M} \frac{\partial H}{\partial \mathbf{q}}. \tag{4.23}
\]
It then follows from equation (4.17) that
\[
\mathbf{q}' = J \frac{\partial H}{\partial \mathbf{q}'}, \tag{4.24}
\]
i.e. the transformation is canonical. In order to prove the necessity of the symplectic condition, we consider infinitesimal canonical transformations.

---

9 A tilde \( \sim \) over a matrix denotes the transpose.

10 \( Sp_n(R) \) is a group since each \( M \) must be invertible for it to be a transformation and because if \( M \) and \( N \) are symplectic, \( (MN)J(MN) = MNJ = MNJ = MNJ = J \).
Since any transformation can be considered to be infinitesimally generated, it will useful to consider infinitesimal canonical transformations. If an infinitesimal canonical transformation satisfies the symplectic condition, it follows that any canonical transformation will satisfy the symplectic condition. Thus, we seek to show that infinitesimal canonical transformations satisfy the symplectic condition to complete the proof of theorem 4.1. An infinitesimal canonical transformation is one in which the phase space coordinates change by an infinitesimal amount:

$$q' = q + \delta q.$$ (4.25)

Such a transformation will differ only infinitesimally from the identity, so to determine an expression for $\delta q$, we seek a generating function for a canonical transformation that differs infinitesimally from the identity. Since $F = q_i'p_i' - q_ip_i$ is the generating function of the identity (see appendix B for a proof), for arbitrary $G(q,p',t)$ and $\epsilon \in \Delta$, letting

$$E = q_ip_i + G\epsilon$$ (4.26)

will yield a canonical transformation that differs only infinitesimally from the identity. Since it is the function $G$ that generates an infinitesimal canonical transformation (ICT), we will refer to $G$ as the generating function for an ICT.

Let us consider an arbitrary generating function $G$. By equation (B.6) we have that

$$q_i' = q_i + \frac{\partial G}{\partial p_i}\epsilon, \quad p_i' = p_i - \frac{\partial G}{\partial q_i}\epsilon.$$ (4.27)

Since $p_i'$ differs only infinitesimally from $p_i$, $\frac{\partial G}{\partial p_i}$ can differ only infinitesimally from $\frac{\partial G}{\partial p_i}$, \(^{11}\) which entails that we may write

$$\delta q_i = \frac{\partial G}{\partial p_i}\epsilon, \quad \delta p_i = -\frac{\partial G}{\partial q_i}\epsilon,$$ (4.28)

which yields an expression for $\delta q$:

$$\delta q = \epsilon J\frac{\partial G}{\partial q}.$$ (4.29)

Since the canonical transformation matrix $M$ is the Jacobian of the canonical transformation, we have that

$$M = \frac{\partial}{\partial q} (q + \delta q) = 1 + \epsilon J\frac{\partial^2 G}{\partial q\partial q}.$$ (4.30)

Since $\frac{\partial^2 G}{\partial q\partial q}$ is symmetric by equality of second order partial derivatives and $J$ is antisymmetric,

$$\tilde{M} = 1 - \epsilon \frac{\partial^2 G}{\partial q\partial q}J,$$ (4.31)

\(^{11}\)This follows because the basis vector corresponding to $p_i'$ can only differ infinitesimally from the basis vector $e_{i+n}$ corresponding to $p_i$, so that

$$\frac{\partial G}{\partial p_i'} = \partial_{e_{i+n} + \epsilon \delta} G = \partial_{e_{i+n}} G + \epsilon \partial_{e} G = \frac{\partial G}{\partial p_i} + \epsilon \partial_{e} G,$$

by the linearity of the directional derivative.
from which it follows that
\[
MJ\tilde{M} = \left(1 + \epsilon J \frac{\partial^2 G}{\partial q \partial \tilde{q}}\right) J \left(1 - \epsilon \frac{\partial^2 G}{\partial \tilde{q} \partial q}\right) J \quad (4.32)
\]
\[
= J + \epsilon J \frac{\partial^2 G}{\partial \tilde{q} \partial q} J - \epsilon J \frac{\partial^2 G}{\partial q \partial \tilde{q}} J \quad (4.33)
\]
\[
= J. \quad (4.34)
\]
Therefore, all ICTs, and hence all canonical transformations, satisfy the symplectic condition. This completes the proof of theorem 4.1 □

Aside from Hamilton’s equations, there are other canonical invariants that are of theoretical interest. One of these is the Poisson bracket. The *Poisson bracket* of two functions \(u\) and \(v\) with respect to the canonical variables \(q\) and \(p\) is defined to be
\[
[u, v]_{q,p} = \frac{\partial u}{\partial q} \frac{\partial v}{\partial p} - \frac{\partial u}{\partial p} \frac{\partial v}{\partial q}. \quad (4.35)
\]
Written in symplectic form we have
\[
[u, v]_{q} = J\frac{\partial u}{\partial Q} J \frac{\partial v}{\partial Q}. \quad (4.36)
\]

**Proposition 4.1** The Poisson bracket is a canonical invariant, i.e.
\[
[u, v]_{q} = [u, v]_{q'}
\]
for any canonical variables \(q\) and \(q'\).

**Proof:** (Goldstein [4]) We observe that
\[
\frac{\partial v}{\partial q} = \tilde{M} \frac{\partial v}{\partial \tilde{q}} \quad (4.37)
\]
and that
\[
\frac{\partial u}{\partial q} = \tilde{M} \frac{\partial u}{\partial \tilde{q}} = \tilde{M} \frac{\partial u}{\partial q} M. \quad (4.38)
\]
Then it follows that we have that
\[
[u, v]_{q} = J\frac{\partial u}{\partial \tilde{q}} J \frac{\partial v}{\partial \tilde{q}} \quad (4.39)
\]
\[
= \tilde{M} J \frac{\partial u}{\partial \tilde{q}} J \frac{\partial v}{\partial \tilde{q}} \quad (4.40)
\]
\[
= \frac{\partial u}{\partial q'} J \frac{\partial v}{\partial q'} = [u, v]_{q'}, \quad (4.41)
\]
since \(M\) is a canonical transformation matrix. □
It is interesting to note that the Poisson bracket is a skew-symmetric bilinear form that satisfies the Jacobi identity, so it is a Lie bracket. This illustrates that the functions on phase space have the structure of a Lie algebra. Since the generating functions $G$ are in this Lie algebra, we begin to see evidence of an analogous relation to the notion of the elements of the Lie algebra as generators of paths in a Lie group. Actually, the analogous structure would be that of a ‘representation’ of a path in a Lie group acting on an Euler space, in this case phase space. The situation here, however, is more complicated than simply this since the structure that bears formal similarity is the generation of paths in phase space by infinitesimal canonical transformations. The generators of infinitesimal shifts in phase space, the analogue of a Lie algebra element, are the the functions $J\frac{\partial G}{\partial q}$ not a vector formed from the function $G$ itself. The other aspect that makes it more complicated is that the linear Lie group that appears here, $Sp_n(\mathbb{R})$ does not transform the canonical coordinates but rather, their time derivatives. Thus, we will not consider the theory developed in the previous section here, but it will be appropriate to do so in considerations of quantum mechanics.

The Lie bracket can be used to develop an alternate formulation of Hamiltonian mechanics. This formulation is based on the fact that Hamilton’s equations can be expressed in terms of the Poisson bracket. For an arbitrary function $u(q, p, t)$, from the chain rule and Hamilton’s equations we have that

$$\frac{du}{dt} = \frac{\partial u}{\partial q} \dot{q} + \frac{\partial u}{\partial p} \dot{p} + \frac{\partial u}{\partial t} = \frac{\partial u}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial u}{\partial p} \frac{\partial H}{\partial q} + \frac{\partial u}{\partial t}. \quad (4.42)$$

We may, however, write this in the form

$$\frac{du}{dt} = [u, H] + \frac{\partial u}{\partial t}, \quad (4.43)$$

which is a generalized formulation of Hamilton’s equations in the Poisson bracket formulation. By substituting each of the canonical variables into equation (4.43) it is easily seen that

$$\dot{q} = [q, H], \quad (4.44)$$

but from the definition of the Poisson bracket we have that

$$[q, H] = J\frac{\partial H}{\partial q}, \quad (4.45)$$

which is seen to be equivalent to Hamilton’s equations and a special case of equation (4.43). It follows immediately from equation (4.43) that if $u$ is a constant of the motion that

$$[H, u] = \frac{\partial u}{\partial t} \quad (4.46)$$

and so if a given function does not depend explicitly on time, then if

$$[H, u] = 0 \quad (4.47)$$

entails that $u$ is a constant of the motion.

We may also conveniently formulate ICTs in this framework. If we consider the ICT $q' = q + \delta q$, then from equation (4.29) and the definition of the Poisson bracket it follows that

$$\delta q = \epsilon[q, G]. \quad (4.48)$$
This provides a convenient framework to analyze the case where \( G = H \). If we also let \( \epsilon = dt \), an infinitesimal time shift, then we have that
\[
\delta q = dt[q, H] = \dot{q} dt = d\dot{q}.
\]
Thus, the change in the phase space coordinates of a system in an infinitesimal time shift is generated by the Hamiltonian. Since the path in phase space is determined by its infinitesimal properties, it follows that the Hamiltonian is the generator of the time evolution of the system.

We will finish this subsection by considering two other kinds of generating functions. Suppose that \( G = p_i \), i.e. the generating function is the \( i \)-th momentum coordinate. It follows from equations (4.28) that
\[
q_i = \delta_{ij} \epsilon, \quad p_i = 0,
\]
so that the \( i \)-th momentum coordinate generates an infinitesimal translation in the \( q_i \) direction. Similarly, the \( i \)-th position coordinate \( q_i \) generates a translation in the negative \( p_i \) direction. To determine the generator of rotations, we consider an infinitesimal rotation. Since equation (4.48) is canonically invariant, \( \delta \mathbf{q} \) can be evaluated in any set of coordinates and retain the same physical significance. Thus, we will use Cartesian coordinates and restrict attention to rotations about the \( z \)-axis. If we let the infinitesimal parameter \( \epsilon = \, d\theta \), an infinitesimal change in angle, then for the \( i \)-th particle we must have that
\[
\delta x_i = -y_i \, d\theta, \quad \delta y_i = x_i \, d\theta, \quad \delta z_i = 0.
\]
Given equations (4.28), this entails that the generating function be of the form
\[
G = x_i p_{i y} - y_i p_{i x},
\]
which is the \( z \)-component of the angular momentum \( \mathbf{L} \). Since the choice of the \( z \)-coordinate is arbitrary, it follows that the generator for a rotation about an axis pointing in the \( \mathbf{n} \) direction is
\[
G = \mathbf{L} \cdot \mathbf{n}.
\]

4.2 Quantum Mechanics: Dynamics and Commutation Relations

One of the historically most important uses of Hamiltonian mechanics was in the development of quantum mechanics. A particularly important tool in this development is the Poisson bracket formulation, since it is very useful for determining the form of the quantum Hamiltonian operator. This becomes possible because the formal structure of quantum mechanics and the Poisson bracket formulation are so similar. One such similarity is that the quantum observables are represented by Hermitian operators (or matrices for finite dimensional state spaces) and the commutator \([A, B] = AB - BA\) of two matrices has the properties of a Lie bracket. The formal relation between quantum mechanics and the classical Poisson bracket formulation is given formal expression as the correspondence principle
\[
[u, v] \rightarrow \frac{1}{i\hbar} [U, V],
\]
where \( u \) and \( v \) are real functions (classical observables) and \( U \) and \( V \) are the corresponding quantum operators. The correspondence principle is not totally satisfactory, however, since, for example, there are quantum observables that have no classical analogue, most
notably spin. Thus, rather than accept the correspondence principle, we will borrow
the infinitesimal generator formalism from classical Hamiltonian mechanics in order to
develop the quantum dynamics and the fundamental quantum commutation relations.
We will see that on several occasions, but by no means all, the correspondence principle
is applicable.

In our treatment of quantum dynamics we will only formally consider finite dimen-
sional state spaces, i.e. $\mathbb{C}^n$. States $|\alpha\rangle$ are represented by rays in $\mathbb{C}^n$ or by the unit
norm representative of a given ray. Experimental setups, or quantum `observables,’ are
represented by Hermitian operators. Given a particular observable $A$ and a state $|\alpha\rangle$,
the state can be expanded in terms of the eigenvectors $|a_i\rangle$ (with eigenvalue $a_i$) of $A$

$$|\alpha\rangle = \sum_i c_i|a_i\rangle. \tag{4.55}$$

The eigenvectors of $A$ represent the possible states of a system subsequent to measure-
ment and the modulus squared of the coefficient of the expansion

$$P_{a_i}(\alpha) = |c_i|^2 \tag{4.56}$$

represents the probability of measuring the system to be in the state $|a_i\rangle$ given that
the system is in the state $|\alpha\rangle$. The measurement process is manifestly not deterministic
and not probability conserving. The ordinary quantum evolution, however, is required
to be deterministic and probability conserving. Probability conservation is achieved
by evolving the state using unitary operators $U$, since the probability information is
contained in the length of the state vector and unitary operators preserve length.

To specify the evolution of the quantum state under some kind of operation, it is
necessary to determine some path in $U_n$ that corresponds the the operation. The kinds
of operations that we are concerned with are motions in some space $\mathbb{R}^m$, be it time ($\mathbb{R}$),
3-space ($\mathbb{R}^3$), Minkowski spacetime ($\mathbb{R}^4$) or phase space ($\mathbb{R}^{2n}$), or rotations in some space
represented by a compact linear Lie group, such as rotations in space $SO_3(\mathbb{R})$ or Lorentz
transformations $SO_{3,1}(\mathbb{R})$. Since the motions or rotations are represented by translations
in some Lie group, we seek to determine the path in $U_n$ that corresponds to a given path
in a Lie group $G$:

$$g(t) \in G \leftrightarrow U(t) \in U_n.$$ 

The reason that we approach the problem this way and do not attempt to determine
the path in $U_n$ given a path in $G$ is that we will consider generators of paths in $U_n$ that correspond to translations in the group $G$ by borrowing notions from classical mechanics.
Once a path in $U_n$ corresponding to a motion in $G$ has been specified with a given initial
state, we may determine the quantum state corresponding to any translation in $G$. For
additive Lie groups, for example, if $|\alpha, g_0\rangle$ is the quantum state corresponding to some
initial position $g(0) = g_0 \in G$, then, $|\alpha, g(t); g_0\rangle = U(t)|\alpha, g_0\rangle$ is the state corresponding
to the final position $g(t)$.

The most natural way to accomplish such a correspondence is to compose a unitary
Lie group representation $\rho: G \rightarrow U_n$, $g_1 * g_2 \mapsto \rho(g_2)\rho(g_1)$, where $*$ represents the group
operation, with a homomorphism $j: J \rightarrow G$ in $G$, $g(t + s) = g(t) * g(s)$. The composition
$U \equiv \rho \circ j$ is then a path in $U_n$ with the desired properties. Since $\rho$ is a representation, $U$
has the properties that we desire of a translation operator. The representation ensures
that the group properties of \( G \) are preserved by the operators \( U(t) \) so that the algebraic relations among elements of \( G \) are satisfied by their representatives in \( U_n \). The representation also preserves the identity \( \iota \) (0 for additive groups, 1 for multiplicative groups) so that

\[
U(t_0) = \rho(g(0)) = \rho(\iota) = 1,
\]

where we set \( t_0 = 0 \). We also have that

\[
U(\epsilon) = \rho(g(\epsilon)) = \rho(\iota \ast \epsilon) = \rho(h\epsilon) = 1 + \xi \epsilon,
\]

where \( h \in g \) and \( \xi \in u_n \). To determine which unitary operators are in \( u_n \) we simply force \( U(\epsilon) \in U_n \), i.e. we require that \( U(\epsilon)U(\epsilon)\dagger = 1 \). This entails that

\[
(1 + \xi \epsilon)(1 + \xi \epsilon)\dagger = (1 + \xi \epsilon)(1 + \xi \epsilon) = 1 + (\xi + \xi \dagger) \epsilon = 1,
\]

which is satisfied if \( \xi = -iG \) with \( G \) Hermitian:

\[
U(\epsilon) = 1 - iG \epsilon, \quad G = G\dagger.
\]

An important thing to notice about this formulation is that

\[
U(t + \epsilon) = \rho(g(t + \epsilon)) = \rho(h\epsilon)\rho(g(t)) = (1 - iG \epsilon)U(t),
\]

so the velocity is constant, which entails that in this case the evolution operator does not depend on the position in \( G \).

In the more general case where the operator \( U \) is to depend on the position, we need to keep track to two group elements, the starting and ending group element for a given translation. Since the only case we consider where this is an issue is time translation, we need not formulate this case in general. Since the time translation operator transforms a quantum state from one time to another as

\[
|\alpha, t_0; t\rangle = U(t_0, t)|\alpha, t_0\rangle,
\]

we want distinct time translation operations to compose, i.e. we require that

\[
U(t_0, t_2) = U(t_1, t_2)U(t_0, t_1), \quad t_0 \leq t_1 \leq t_2.
\]

This entails that \( U(t, t) = 1 \) and, hence, that

\[
U(t, t + \epsilon) = 1 + \xi(t) \epsilon,
\]

\( \xi(t) \in u_n \). Therefore, we have that

\[
U(t, t + \epsilon) = 1 - iG(t) \epsilon, \quad G(t) = G(t)\dagger,
\]

where \( G(t) \) is the generating function of the time evolution.

Borrowing from classical mechanics the notion that the Hamiltonian is the generator of time evolution, we obtain

\[
U(t, t + \epsilon) = 1 - i \left( \frac{H}{\hbar} \right) \epsilon,
\]

22
where $H$ is the quantum Hamiltonian of the system under consideration and $\hbar$, the natural quantum unit of action, is included to make the generator dimensionless. Now we consider the operation $U(t_0, t + \epsilon)$. Letting $dt = \epsilon$ we have that

$$U(t_0, t + dt) = U(t, t + dt)U(t_0, t)$$  \hspace{1cm} (4.67)

$$= \left(1 - i \left(\frac{H}{\hbar}\right) dt\right)U(t_0, t)$$  \hspace{1cm} (4.68)

$$= U(t_0, t) - i \left(\frac{H}{\hbar}\right) U(t_0, t) dt.$$  \hspace{1cm} (4.69)

By fixing $t_0$, $U(t_0, t + dt)$ is a function from $\mathbb{R}$ to $U_n$, so we have that (since $U_n$ is Euclidean)

$$U(t_0, t + dt) = U(t_0, t) + \frac{\partial U}{\partial t}(t_0, t) dt.$$  \hspace{1cm} (4.70)

From equations (4.69) and (4.70) it follows that

$$\frac{\partial U}{\partial t}(t_0, t) dt = -i \left(\frac{H}{\hbar}\right) U(t_0, t) dt,$$  \hspace{1cm} (4.71)

which entails that

$$i\hbar \frac{\partial U}{\partial t}(t_0, t) = HU(t_0, t),$$  \hspace{1cm} (4.72)

by PMC. By applying this to the state vector, we obtain the familiar Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\alpha, t_0; t\rangle = H|\alpha, t_0; t\rangle.$$  \hspace{1cm} (4.73)

In the above discussion, we have assumed that the time evolution operator acts on the state vector. This is called the Schrödinger picture. If, on the other hand, we consider the time evolution operator to act on observables, then we obtain an alternate formulation called the Heisenberg picture. To make the translation, if $A^S$ is an observable in the Schrödinger picture, then $A^H = U^\dagger A^S U$ is the corresponding observable in the Heisenberg picture. Let us consider cases where $A^S$ does not depend explicitly on time. Then, letting $U(t) \equiv U(t, t_0)$ we have that

$$\frac{dA^H}{dt} = \frac{\partial U^\dagger}{\partial t}(t) A^S U(t) + U^\dagger(t) A^S \frac{\partial U}{\partial t}(t)$$  \hspace{1cm} (4.74)

$$= \frac{1}{i\hbar} (-U^\dagger(t)HAU(t) + U^\dagger(t)AHU(t))$$  \hspace{1cm} (4.75)

$$= \frac{1}{i\hbar} (-U^\dagger(t)HU(t)U^\dagger(t)AU(t) + U^\dagger(t)AU(t)U^\dagger(t)HU(t))$$  \hspace{1cm} (4.76)

$$= \frac{1}{i\hbar} (AH^H - H^H A^H),$$  \hspace{1cm} (4.77)

where we have let $H^H \equiv U^\dagger(t)HU(t)$. We may rewrite this last equation, omitting the superscript $H$, as

$$\frac{dA}{dt} = \frac{1}{i\hbar} [A, H],$$  \hspace{1cm} (4.78)

which exactly what is obtained from equation (4.43) in this case and by applying the correspondence principle.
Now we will turn to the consideration of spatial translations. In order to treat this problem I will take certain liberties with the mathematics. I have not developed the theory of SIA enough to treat functionals and operators on infinite dimensional Hilbert spaces\(^{12}\), so I will simply assume that they exist in \(S\) and, since they exist in \(S\), have the requisite smoothness properties. We are only interested in infinitesimal translations here, so we may use the general theory developed above, where the Lie group that we are considering translations in is \(\mathbb{R}^3\). We denote the infinitesimal spatial translation operator by \(T(dx)\), where \(dx\) is a microvector, such that if \(|x\rangle\) represents a quantum state localized at the position \(x\), then the action of the unitary operator \(T\) is such that
\[
T(dx)|x\rangle = |x + dx\rangle. \tag{4.79}
\]
If we take from classical mechanics the notion that momentum is the generator of space translations, then by the earlier general considerations we have for a translation along a given, say the \(x\)-, axis
\[
T(dx) = 1 - \frac{i}{\hbar} p_x dx, \tag{4.80}
\]
where \(\hbar\) is included for dimensional considerations. Then, it follows naturally that, for a general infinitesimal shift, we have
\[
T(dx) = 1 - \frac{i}{\hbar} \mathbf{p} \cdot dx \tag{4.81}
\]
Let \(\hat{x}\) be the position operator, i.e. \(\hat{x}|x\rangle = x|x\rangle\).

**Proposition 4.2** The position and momentum operators satisfy the fundamental canonical commutation relations, i.e.
\[
[x_i, p_i] = i\hbar \delta_{ij}, \tag{4.82}
\]

**Proof:** (essentially Sakurai [12]) From the properties of the operator \(T\) we have that
\[
\hat{x}T(dx)|x\rangle = \hat{x}|x + dx\rangle = (x + dx)|x + dx\rangle, \tag{4.83}
\]
and that
\[
T(dx)\hat{x}|x\rangle = T(dx)x|x\rangle = x|x + dx\rangle. \tag{4.84}
\]
These two equations then entail that
\[
\hat{x}T(dx)|x\rangle - T(dx)\hat{x}|x\rangle = dx|x + dx\rangle = dx|x\rangle, \tag{4.85}
\]
where the last equality follows because we must have \(|x + dx\rangle = |x\rangle + v dx\) for some \(v\) and because \(dx\) is nilpotent. Thus, we have that
\[
[x, T(dx)] = dx. \tag{4.86}
\]
Substituting our expression for \(T(dx)\) it follows that
\[
\hat{x}\mathbf{p} \cdot dx - \mathbf{p} \cdot d\hat{x} = i\hbar dx. \tag{4.87}
\]
Now, if \(\hat{i}\) is a unit vector along the \(i\) axis, where \(i\) ranges from 1 to 3, by letting \(dx = \hat{i}\epsilon\) and forming the dot product of both sides of equation (4.87) with \(\hat{j}\), it follows that
\[
(x_j p_i - p_i x_j)\epsilon = i\hbar (\hat{i} \cdot \hat{j})\epsilon \tag{4.88}
\]
so that by PMC
\[
[x_i, p_i] = i\hbar \delta_{ij}. \tag{4.89}
\]
\(^{12}\) Such spaces would be the SIA equivalents of spaces such as the countable \(l^2\) or the uncountable \(L^2\). Such spaces would presumably be some exponential object, such as \(C^\infty\), but the consideration of the theory of these objects is beyond the scope of this paper.
We will finish with a consideration of the case of rotations. Since we are only interested in infinitesimal rotations here, we may apply the general framework developed above to this case, where the Lie group here is \( SO_3(\mathbb{R}) \). However, things are more involved than in the case of infinitesimal translations. Working in the group \( SO_3(\mathbb{R}) \) we observe that the nilsquare infinitesimals are not sufficient for an analysis of rotations. We may see this explicitly, however, since for a nilsquare infinitesimal \( \epsilon \), the infinitesimal rotation matrix for about the \( x \)-axis (see appendix C) is

\[
R_x(\epsilon) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - \epsilon & 0 \\ 0 & \epsilon & 1 \end{pmatrix}
\]  

(4.90)

and the infinitesimal rotation matrix about the \( y \)-axis is

\[
R_y(\epsilon) = \begin{pmatrix} 1 & 0 & \epsilon \\ 0 & 1 & 0 \\ -\epsilon & 0 & 1 \end{pmatrix}
\]  

(4.91)

which entails that

\[
R_x(\epsilon)R_y(\epsilon) = \begin{pmatrix} 1 & 0 & \epsilon \\ 0 & 1 - \epsilon & 0 \\ -\epsilon & 0 & 1 \end{pmatrix} = R_y(\epsilon)R_x(\epsilon).
\]  

(4.92)

Thus, using nilsquare infinitesimals rotations about different axes commute, which we know is not the case for finite rotations. Intuitively this occurs because nilsquare infinitesimals are too small to analyze curvature and the non-commutativity of rotations in 3-space are as a result of the curvature of the sphere. Since the non-commutativity is crucial for our analysis, we will consider the larger nilcube infinitesimals.

Now we may treat the problem of rotations in quantum mechanics. Following the lead of classical mechanics and the above general theory we will take the generator \( G \) of an infinitesimal translation to be of the form \( \frac{1}{\hbar} J \cdot \hat{n} \), where \( J \) is the quantum angular momentum operator, giving

\[
\mathcal{R}(\hat{n}, d\theta) = 1 - \frac{i}{\hbar} J \cdot \hat{n} d\theta,
\]  

(4.93)

where we take \( d\theta \) to be a nilcube infinitesimal. In appendix C we established that

\[
R_x(d\theta)R_y(d\theta) - R_y(d\theta)R_x(d\theta) = R_z(d\theta^2) - 1.
\]  

(4.94)

Since \( \mathcal{R} \) is a representation,\(^{13}\) it follows that

\[
\mathcal{R}(\hat{x}, d\theta)\mathcal{R}(\hat{y}, d\theta) - \mathcal{R}(\hat{y}, d\theta)\mathcal{R}(\hat{x}, d\theta) = \mathcal{R}(\hat{z}, d\theta^2) - 1.
\]  

(4.95)

By substituting our expression for \( \mathcal{R} \) we have

\[
\left( 1 - \frac{i}{\hbar} J_x d\theta \right) \left( 1 - \frac{i}{\hbar} J_y d\theta \right) - \left( 1 - \frac{i}{\hbar} J_y d\theta \right) \left( 1 - \frac{i}{\hbar} J_x d\theta \right) = 1 - \frac{i}{\hbar} J_x d\theta^2 - 1,
\]  

(4.96)

so that

\[
1 - \frac{i}{\hbar} J_x d\theta - \frac{i}{\hbar} J_y d\theta - \frac{1}{\hbar^2} J_x J_y d\theta^2 - 1 + \frac{i}{\hbar} J_y d\theta + \frac{i}{\hbar} J_x d\theta + \frac{1}{\hbar^2} J_y J_x d\theta^2 = -\frac{i}{\hbar} J_z d\theta^2.
\]  

(4.97)

\(^{13}\)It is essentially a representation since the fact that we are only considering infinitesimal translations makes the considerations of paths superfluous.
Thus, we have that
\[ \frac{1}{\hbar^2} (J_x J_y - J_y J_x) d\theta^2 = \frac{i}{\hbar} J_z d\theta^2, \]
and so by PMC it follows that
\[ [J_x, J_y] = i\hbar J_z. \]
Because equation (4.94) holds cyclically, it follows that
\[ [J_i, J_j] = i\hbar \epsilon_{ijk} J_k, \]
which are the fundamental commutation relations of angular momentum.

5 Conclusion

We have seen that smooth infinitesimal analysis provides a natural framework for formulating rigorous versions of the intuitive proofs used in the basic theory of the classical Lie groups and in the treatment of dynamics in classical and quantum mechanics. In almost all cases, the rigorous SIA versions follow very closely the informal reasoning used by physicists. Indeed, SIA should apply similarly to the treatment of variational methods and to other physical theories such as special and general relativity and statistical mechanics. It would take a series of additional projects to examine these areas.

Despite how natural the application of SIA to dynamics is, there is one particular aspect that is not naturally accommodated. This problem occurs in the consideration of paths in matrix spaces. One way of stating the limitation is the difficulty in providing a straightforward proof of the following proposition:

**Proposition 5.1** A path \( X(t) \) is a one-parameter subgroup iff \( X(0) = 1 \) and the velocity is constant.

The forward direction is simple, but there does not seem to be a ready proof to the reverse direction. Classically, this is naturally accomplished by the consideration of the exponential matrix \( e^{At} \), where \( A = X'(0) \) and
\[ e^{At} = \sum_{n=1}^{\infty} \frac{(At)^n}{n!}, \]
from which we can show that
\[ e^{A(t+s)} = e^{At} e^{As}. \]
SIA, however, does not have limits and so this definition of the exponential matrix is not available. As is easily seen, the generation of solutions to \( \dot{X} = AX, \quad X(0) = 1 \), using canonical forms does not lend itself to a definition of an exponential matrix, even using the nicer Jordan form considered in appendix A.

It is possible that the consideration of general Lie theory may provide a framework in which one-parameter subgroups are more naturally characterized,\(^\text{14}\) but there does not seem to be a ready solution to this predicament here. Besides the fact that the lack of a SIA equivalent of the exponential matrix prevents a clear way of characterizing one-parameter subgroups, there is also the fact that the exponential matrix is the natural

\(^{14}\) It is always, of course, possible that a natural natural proof has escaped my notice or my ingenuity.
way of presenting many of the evolution operators in quantum mechanics. Without the exponential matrix, the SIA version of quantum dynamics is not likely to be of much interest to physicists. It would, of course, be possible to use nonstandard analysis to treat quantum mechanics, since it has limits available in the formal theory, but then one must sacrifice nilpotent infinitesimals, which it not a satisfactory exchange. Thus, I conclude, that a major obstacle to the ability to formulate dynamical theories within SIA is the development of an equivalent of the classical exponential matrix.
References


A Existence and uniqueness using Jordan canonical form

We provide an alternative, but less general, proof of the existence of a unique solution to equation (3.19): 
\[ \dot{X} = AX, \quad X(0) = 1, \]

using the Jordan canonical form. The proof begins in the same way as the proof using the rational canonical form. We observe that 
\[ \frac{dX_i}{dt} = \frac{dX_j}{dt} = A_kX_j^k, \]

so that the \( j \)-th column of \( \frac{dX}{dt} \) depends only on the \( j \)-th column of \( AX \). Thus, equation (3.19) is equivalent to \( n \) equations of the form 
\[ \dot{x}_i = Ax_i, \quad [x_i(0)]^j = \delta_i^j \]

(A.1)

where \( x_i \in \mathbb{C}^n \) is a column vector. Rather than consider the rational canonical form, we use the fact that every \( A \in \mathbb{C}^{n \times n} \) for which the characteristic polynomial splits can be expressed in the form 
\[ A = QJQ^{-1}, \]

(A.2)

where \( J \) is the Jordan canonical form for \( A \) and \( Q \) is an invertible matrix. In general, \( J \) is of the form 
\[ J = \begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_k, \end{pmatrix} \]

(A.3)

where each \( O \) is a zero matrix of the appropriate size, and each \( A_j \) is a Jordan block, \( i.e. \) of the form 
\[ A_j = \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_k \\ 0 & 0 & 0 & \cdots & \lambda_k \end{pmatrix} \]

(A.4)

We may now note the following two things. Firstly, if we let \( \bar{x}_i \equiv Q^{-1}x^i \), then if \( \bar{x}_i \) is a solution of 
\[ \frac{d\bar{x}_i}{dt} = J\bar{x}_i, \quad [\bar{x}_i(0)]^j = [Q^{-1}]^{ji}, \]

(A.5)

then \( x_i = Q\bar{x}_i \) is a solution of equation (3.20). Thus, we may restrict our attention to the case where \( A \) is a Jordan matrix. Secondly, we observe that if \( A \) is a Jordan matrix, equation (3.20) splits into \( k \) independent equations of the form 
\[ \frac{dx_{ik}}{dt} = A_kx_{ik}, \quad [x_{ik}(0)]^j = a^j, \]

(A.6)

for some \( a^j \). Thus, the general problem of solutions to equation (3.19) reduces to solutions of equation (A.6).

We will demonstrate existence and uniqueness of solutions to equation (A.6) by explicit construction. Suppose that \( A_k \) is a matrix of dimension \( l \), then we seek \( l \) dimensional column vector \( x_{ik} \) as the solution. To clarify notation, we will let \( x \equiv x_{ik} \). Since \( A_k \) is a Jordan block matrix, we obtain from equation (A.6) the \( l \) equations 
\[ (x^j)' = \lambda_k x^j + x^{j+1}, \quad j < l \]

(A.7)

\[ (x^l)' = \lambda_k x^l, \]

(A.8)
with the initial conditions \( x_j(0) = a_j \). \( x^j = e^{\lambda k t} + a^j \) is a solution to equation (A.8) and by IP it is unique. We may then find solutions to equations (3.26) by induction. Suppose that there is a unique \( x_{j+1} \) such that \( x_{j+1} = a^{j+1} \). By (A.7) we know that

\[
(x_j)' - \lambda_k x_j = x_{j+1}.
\] (A.9)

Multiplying both sides by \( e^{-\lambda_k t} \) we may obtain

\[
(x_j e^{-\lambda_k t})' = e^{-\lambda_k t} x_{j+1}.
\] (A.10)

Now, by QCIP there is a unique function \( f = x_j e^{-\lambda_k t} \) such that \( f(0) = a_j \) and \( f' = e^{-\lambda_k t} x_{j+1} \). Stated otherwise, we have that

\[
x_j e^{-\lambda_k t} = g(t) + c,
\] (A.11)

where \( c \) is a constant of integration and \( g(t) = \int e^{-\lambda_k t} x_{j+1} dt \) is unique. Therefore, we have that

\[
x_j = e^{\lambda_k t} (g(t) + c).
\] (A.12)

Since we require that \( x_j(0) = a_j \), we obtain a unique value for \( c \)

\[
c = a_j - g(0),
\] (A.13)

which entails that \( x_j \) is unique. Thus, by induction, there exists a unique solution to the initial value problem (A.6). It, therefore, follows that there exists a unique solution to equation (3.19) satisfying the initial condition.\(^{15}\)

### B Identity Canonical Transformation

In this section, we prove that the generating function

\[
F = q_i p_i' - q_i' p_i,
\] (B.1)

is the generating function of the identity transformation. Consider the generating function

\[
F = E(q, p', t) - q_i' p_i'.
\] (B.2)

Then we have that

\[
\frac{dF}{dt} = \frac{dE}{dt} - q_i' p_i' - p_i' q_i',
\] (B.3)

so that by substituting this in equation (4.15) we obtain

\[
p_i q_i - H = -q_i' p_i' - K + \frac{dE}{dt}.
\] (B.4)

By the chain rule we have

\[
\frac{dE}{dt} = \frac{\partial E}{\partial q_i} q_i + \frac{\partial E}{\partial p_i} p_i' + \frac{\partial E}{\partial t},
\] (B.5)

so that from equation (B.4) we obtain

\[
p_i = \frac{\partial E}{\partial q_i}, \quad q_i' = \frac{\partial E}{\partial p_i'}, \quad K = H + \frac{\partial E}{\partial t}.
\] (B.6)

Now, if we let \( E = q_i p_i' \), we have that

\[
q_i' = q_i, \quad p_i' = p_i,
\] (B.7)

so that \( F = E - q_i' p_i' \) is the identity transformation.

\(^{15}\)Note that this construction only works for matrices that have a Jordan canonical form.
C Infinitesimal Rotations

In this section we will determine the matrices for infinitesimal rotations in $\mathbb{R}^3$, i.e., the infinitesimal elements of $SO_3(\mathbb{R})$. We will not do this in general but simply for infinitesimal rotations about the $x$-, $y$- and $z$-axes, from which any arbitrary infinitesimal rotation matrix can be obtained. The rotation matrix for a rotation of angle $\theta$ about the $z$-axis is

$$ R(\hat{z}, \theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (C.1) $$

Similarly, for rotations about the $x$- and $y$-axes we have that

$$ R(\hat{x}, \theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta - \sin \theta & 0 \\ 0 & \sin \theta & \cos \theta \end{pmatrix}, \quad R(\hat{y}, \theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}. \quad (C.2) $$

We may obtain the corresponding infinitesimal rotation matrices simply by considering a rotation of an infinitesimal angle $d\theta \in \Delta_k$ for some $k$. Since the highest order infinitesimal we are interested in is second-order, we will consider the case where $k = 2$ and the first-order versions of the matrices are obtained when $d\theta^2 = 0$. From Taylor's theorem (Bell [2], p. 92), we have that for $\sin \theta$ and $\cos \theta$

$$ \sin(d\theta) = d\theta, \quad \cos(d\theta) = 1 - d\theta^2. \quad (C.3) $$

It then follows that the matrix for an infinitesimal rotation about the $z$-axis is

$$ R(\hat{z}, d\theta) = \begin{pmatrix} 1 - \frac{d\theta^2}{2} & -d\theta & 0 \\ d\theta & 1 - \frac{d\theta^2}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (C.4) $$

Similarly, for rotations about the $x$- and $y$-axes we have that

$$ R(\hat{x}, d\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - \frac{d\theta^2}{2} & -d\theta \\ 0 & d\theta & 1 - \frac{d\theta^2}{2} \end{pmatrix}, \quad R(\hat{y}, d\theta) = \begin{pmatrix} 1 - \frac{d\theta^2}{2} & 0 & d\theta \\ 0 & 1 & 0 \\ -d\theta & 0 & 1 - \frac{d\theta^2}{2} \end{pmatrix}. \quad (C.5) $$

Now, it is easily verified that

$$ R(\hat{x}, d\theta)R(\hat{y}, d\theta) = \begin{pmatrix} 1 - \frac{d\theta^2}{2} & 0 & d\theta \\ d\theta^2 & 1 - \frac{d\theta^2}{2} & -d\theta \\ -d\theta & d\theta & 1 - d\theta^2 \end{pmatrix}, \quad (C.6) $$

and that

$$ R(\hat{y}, d\theta)R(\hat{x}, d\theta) = \begin{pmatrix} 1 - \frac{d\theta^2}{2} & d\theta^2 & d\theta \\ 0 & 1 - \frac{d\theta^2}{2} & -d\theta \\ -d\theta & d\theta & 1 - d\theta^2 \end{pmatrix}. \quad (C.7) $$

It follows from this that

$$ R(\hat{x}, d\theta)R(\hat{y}, d\theta) - R(\hat{y}, d\theta)R(\hat{x}, d\theta) = \begin{pmatrix} 0 & -d\theta^2 & 0 \\ d\theta^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = R(\hat{z}, d\theta^2) - I. \quad (C.8) $$

The significance of equation (C.9) is that for first-order infinitesimal rotations, $d\theta^2 = 0$, so that rotations about different axes commute, but for second-order infinitesimal rotations, rotations about different axes do not commute, as is the case for finite rotations.