MODELING SYMMETRIES: AN ANALYSIS ON THE INTERDISCIPLINARY APPLICATIONS OF LIE GROUPS

by

Reece Humphreys

A Thesis Submitted to the Faculty of

The Wilkes Honors College

in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Biological and Physical Sciences

with a Concentration in Mathematics

Wilkes Honors College of

Florida Atlantic University

Jupiter, Florida

May 2021

MODELING SYMMETRIES: AN ANALYSIS ON THE INTERDISCIPLINARY APPLICATIONS OF LIE GROUPS

by

Reece Humphreys

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Warren McGovern, and has been approved by the members of their supervisory committee. It was submitted to the faculty of the Harriet L. Wilkes Honors College and was accepted in partial fulfillment of the requirements for the degree of Bachelor of Science in Biological and Physical Sciences.

SUPERVISORY COMMITTEE:

_____

Dr. Warren McGovern

_____

Dr. Terje Hill

_____

Dean Justin Perry, Harriet L. Wilkes Honors College

_____

Date

Acknowledgments

This thesis is dedicated to Dr. Warren McGovern and Dr. Terje Hill. Both of them inspired a passion for mathematics that I could never have imagined and provided immense support throughout my four years at the Honors College. Additionally, I would like to give a special thanks to Dr. Timothy Steigenga, who provided me countless opportunities and showed me there is always a path forward.

Abstract

Author: Reece Humphreys

Title: Modeling Symmetries: An analysis on the

                     interdisciplinary applications of Lie groups

Institution: Harriet L. Wilkes Honors College, Florida Atlantic University

Thesis Advisor: Dr. Warren McGovern

Degree: Bachelor of Science in Biological and Physical Sciences

Concentration: Mathematics

Year: 2021

Symmetries are a driving force in the universe and continue to reveal themselves the deeper we look. While many undergraduate mathematics students are introduced to symmetries with group theory, many practical applications are often overlooked. The application of symmetries provides deep insights into various problems and can frequently simplify complex mathematics. Applying symmetries typically requires high-level mathematics that can be prohibitive for people within other disciplines. As such, this paper explores how Lie groups, a mathematical structure for representing symmetries, can be utilized in computing, physics, and control theory to solve practical problems from within these fields. It also serves as an introduction to the mathematics necessary to utilize Lie groups.

Contents

1 Introduction

1.1 Mathematically modeling symmetries

Groups are algebraic objects that are central to understanding the structure of mathematics. A group is a set combined with a binary operation that satisfies the axioms of associativity, identity, closure, and invariability [13]. This structure is one of the simplest in mathematics, yet encoded within its properties are the beginnings of a crucial idea in mathematics, the notion of symmetries. A Lie group is a special kind of continuous/topological group. It is a group equipped with a topology and has the idea of symmetries built into it. More specifically, it is a smooth manifold in which the group operation is continuous. The definitions and relationships between these mathematical objects will be explored in chapter 2 of this paper. The difficulty with utilizing these powerful algebraic and geometric objects is the rigorous, abstract mathematics required to understand them. As such, Lie groups are under-utilized compared to more intuitive mathematics such as calculus. For this reason, it is the goal of this paper to not only build up an intuitive understanding of the relationships between these mathematical constructs but also show how the properties of these objects can be utilized in an interdisciplinary framing.

1.2 Improving numerical integration and the energy drift problem

One of the newer applications discovered for Lie groups relates to numerical methods/numerical computing. In computing, a typical application of mathematics is the use of integrators to evaluate differential equations.

However, this method often requires compromising either computational speed or accuracy to create an efficient equation. Sacrificing either factor makes it difficult for researchers to make breakthroughs in their fields. By utilizing Lie groups in computation, the symmetries of nature can be encoded into the integrator's design. This can lead to accurate and fast computations that describe the differential equation's essence to be solved more accurately. For example, using symmetries to perform better integrations is useful when studying possible solutions to the energy drift problem. One of the fundamental laws of physics explains that the total energy in a closed system must remain constant throughout time. However, due to the error that often results from performing numerical integration, the total energy of a system tends to drift. This is a significant shortcoming of integrators as the integration results will no longer satisfy a crucial law of physics. The energy drift problem and solutions to it will be explored further in-depth in chapter 3.

1.3   Noether's theorem

Physics is another field that frequently makes use of symmetries to solve problems and gain new insights. A theorem integral to the application of Lie groups in physics is Noether's theorem, which informally states the following ... if a system has a continuous symmetry, then there are corresponding quantities whose values are conserved in time. This provides a direct link between physical conservation and the Lie group structure. Physical conservation is simply something that remains unchanged throughout the evolution of the system. The most well-known example is the conservation of energy, which states that an isolated system's total energy remains

constant over time, i.e., energy is conserved over time. While Noether's theorem plays a significant role in theoretical physics and the calculus of variations, it is not without its constraints. Notably, it does not apply to systems that cannot be modeled with a Lagrangian alone. An application of Noether's theorem is explored in chapter 4.

1.4 Controllability for Affine Non-Linear Systems

The final application studied in this paper is determining the controllability of certain control systems. This falls under the control theory field, which engineers use to analyze how state inputs can move a system to some desired output state. Control theory is a widely used field with applications ranging from aerospace engineering to robotics. One of the first steps in control theory is performing an analysis of a system before deciding the best control strategy. Lie groups play a unique role in affine non-linear control systems that allow us to execute this analysis.

2 Definitions

2.1 Groups

To build up to the definition of a Lie group, first, we must understand its construction from both an algebraic and topological perspective.

The most elementary form of structure used in mathematics is the **set** which is used to specify a collection of objects. These objects can be anything from numbers to letters and symbols. They can also be finite such as the set $\{1, 4, 9\}$ or infinite such as the set of all real numbers. An important property of sets is that the elements contained within them must be unique. For example, if there exists a set containing the letter 'a' five times, it would only have one element, 'a', as 'a' is the only unique element.

Definition 2.1. A **set** is a collection of well defined and distinct objects, which is itself an object.

Example 2.1. $\emptyset = \{\}$ *is a set, a set containing nothing. This set is called the **empty set** and it is unique.*

Example 2.2. $\{\{1, 2\}, \emptyset\}$ *is a set. More specifically, it is a set containing a set and the empty set.*

Given two sets X and Y we can form a new set using the **Cartesian Product**, which is the set of all ordered pairs (x, y) where x is an element in X, and y is an element in Y.

Definition 2.2. The **Cartesian Product** of sets X and Y is denoted $X \times Y$ and is defined using set-builder notation by

$$X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}$$

Example 2.3. $\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$

Recall that a **function** takes some input x and maps it to a corresponding $f(x)$. Note that $f(x)$ does not need to be the same type of number as x. For example, suppose $X = \mathbb{N}$ and $Y = \mathbb{Z}$, we could define a function that maps the natural numbers into the integers as follows: $f(x) = -x$. This would map the number 1, which is a whole number, to -1 , which is an integer. For this reason, it can be useful to think of functions in a more abstract way that tells us information about what the function is mapping to and from. For the prior example we would write $f : \mathbb{N} \to \mathbb{Z}$, where $f(x) = -x$. More abstractly, functions map from one set to another.

Definition 2.3. For two sets, X and Y we define a **function** $f : X \to Y$ which maps an element $x \in X$ to some $y \in Y$. More formally, a function from X to Y is a subset $f \subseteq X \times Y$ which satisfies the following:

1. $\forall x \in X \, \exists y \in Y$ s.t. $(x, y) \in f$
2. $\forall (x, y_1), (x, y_2) \in f \to y_1 = y_2$

Definition 2.4. The **distance function** $d : \mathbb{R}^n \to \mathbb{R}^+$ is defined on $\mathbb{R}^n$ as

$$d(x_1, x_2, ..., x_n) = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} \tag{1}$$

Using the prior definitions we can now define a **binary operation** which is central in the definition of a group. A binary operation is a function that takes two elements from a set, and maps it back into that same set. The commonly used operations of addition, division, multiplication, and subtraction are all examples of binary operations. For example, consider $2 + 3 = 5$, where both 2 and 3 are elements of the naturals and the number 5 again is a natural number. As such, addition is a binary operation on

the natural numbers. It should be noted that subtraction would not be a binary operation on the natural numbers as it is possible to get a value not in the set. For example, $2 - 3 = -1$ is not an element of the natural numbers.

Definition 2.5. Let X be a set. A **binary operation** on X is a function, $* : X \times X \to X$. For any two elements $x_1, x_2 \in X$ we can denote the binary operation using **infix notation** as $x_1 * x_2$ instead of $*((x_1, x_2))$.

A **group** is an algebraic structure that by its construction, encodes important information about symmetries. The definition requires a set equipped with a binary operation to satisfy various axioms such as closure, associativity, and identity. While these properties may seem unassuming at first glance, it provides a foundational object for discussing the structure that exists within mathematics.

Definition 2.6. A **group** is a set X together with a binary operation $*$ on X that satisfies the following conditions

- $\forall x_1, x_2, x_3 \in X$, $x_1 * (x_2 * x_3) = (x_1 * x_2) * x_3$ (**Associativity**)

- $\exists x_i \in X$ such that $\forall x \in X$, $x_i * x = x * x_i = x$ (**Identity**)

- $\forall x \in X \ \exists x^{-1}$ such that $x * x^{-1} = x^{-1} * x = x_i$ (**Inverse**)

It should be noted that the identity of a group, x_i, is unique and that the inverse of an element is also unique.

Example 2.4. $\mathbb{Z}$ *with addition is a group, denoted $(\mathbb{Z}, +)$. The inverse of an element $a \in \mathbb{Z}$ is $-a$ and the identity is 0.*

An important property to note is that **commutativity is not a requirement** for a group. Commutativity on X would be defined as follows: For all $x_1, x_2 \in X$ and a binary operation *, we have $x_1 * x_2 = x_2 * x_1$. A group where the binary operation is commutative is called an **Abelian group**.

Example 2.5. *The set of real numbers, $\mathbb{R}$, with 0 removed, denoted $\mathbb{R}\backslash\{0\}$ or $\mathbb{R}^*$, combined with the operation of multiplication, forms a group, $(\mathbb{R}^*, \bullet, 1)$. One should remove 0 as it does not have a multiplicative inverse. More specifically, $\mathbb{R}^*$ forms an Abelian group.*

Example 2.6. *A **permutation** of a set X is a function $\sigma : X \to X$ that is a bijective map. Permutation maps have inverses by definition of being bijections and are associative under composition of maps. Additionally, there is an identity permutation $\sigma_i : X \to X$ defined by $\forall x \in X$, $\sigma_i(x) \to x$. As such, the set of all permutations on $X = \{x_1, x_2, ..., x_n\}$ forms a group under composition called the **symmetric group** S_n of degree n.*

2.2 Matrices and Matrix Groups

A **matrix** is a rectangular array of numbers, symbols, operations, or expressions that is arranged in rows and columns. Matrices are powerful objects that can be used to simplify linear equations, transform spaces, and are the preferred structure for containing data used in numerical computations.

Every matrix has two numbers associated with it. These two numbers when written as $r \times c$, where r is the number of rows and c is the number of columns, are called the **dimension** of the matrix. For example, the dimension of a matrix containing two rows and two columns would be 2×2.

The following are all examples of matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad (2 \times 3 \text{ matrix containing numbers})$$

$$\mathbf{B} = \begin{pmatrix} x^2 & xy \\ y + x^3 & 8x \end{pmatrix} \qquad (2 \times 2 \text{ matrix containing expressions})$$

$$\mathbf{C} = \begin{pmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \end{pmatrix} \qquad (3 \times 1 \text{ matrix containing operations})$$

To denote matrices, I will be using boldface uppercase letters as seen above. This convention will be maintained throughout the rest of this thesis to distinguish between a set X and a matrix $\mathbf{X}$.

An individual element of a matrix $\mathbf{A}$ is typically denoted $a_{i,j}$ and is called an **entry** of the matrix, where i is index of the row and j is the index of the column. As such, for a $m \times n$ matrix, $1 \le i \le m$ and $1 \le j \le n$. For example, in the above matrix $\mathbf{A}$, $a_{1,2} = 2$, $a_{2,3} = 6$. For matrix $\mathbf{B}$, $b_{2,2} = 8x$, and for matrix $\mathbf{C}$, $c_{3,1} = \frac{\partial}{\partial z}$. This type of notation is called **index notation** and can also be used to define matrices.

2.2.1 Types of Matrices

Definition 2.7. A **row matrix** is a matrix that has only one row, but any number of columns. As such, it would be a $1 \times m$ matrix. For example,

$$\mathbf{A} = \begin{pmatrix} 1 & 5 & -3 & \sqrt{2} & -12 \end{pmatrix}$$

is a row matrix with dimensions 1×5.

Definition 2.8. A **column matrix** is a matrix that has only one column, but any number of rows. As such, it would be a $n \times 1$ matrix. For example,

$$\mathbf{A} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

is a column matrix with dimensions 3×1.

Definition 2.9. A **square matrix** is a matrix that has the same number of rows and columns. As such, it would be a $n \times n$ matrix. For example,

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

is a square matrix with dimensions 3×3.

Definition 2.10. The **main diagonal** of a matrix $\mathbf{A}$ is the collection of entries $a_{i,j}$ where $i = j$.

Definition 2.11. An **identity matrix** is a $n \times n$ square matrix that contains ones on the main diagonal and zeroes everywhere else. It is typically denoted $\mathbf{I_n}$. The identity matrix $\mathbf{I_3}$ is shown below.

$$\mathbf{I_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Definition 2.12. A **diagonal matrix** is a matrix that has zeros in position $a_{i,j}$ when $i \neq j$. For example,

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is a 3×3 square, diagonal matrix.

Definition 2.13. A **triangular matrix** is special type of square matrix. A square matrix is called **upper triangular** if all entries below the main diagonal are zero. More formally, a square matrix is upper triangular if it has the following form:

$$a_{i,j} = \begin{cases} a_{i,j} & \text{for } i \leq j \\ 0 & \text{for } i > j \end{cases}$$

Similarly, a square matrix is lower triangular if it has the following form:

$$a_{i,j} = \begin{cases} a_{i,j} & \text{for } i \geq j \\ 0 & \text{for } i < j \end{cases}$$

2.2.2 Matrix Operations

Definition 2.14. The **scalar multiplication** of a matrix $\mathbf{A}$ with a scalar $\alpha \in \mathbb{R}$ is denoted by $\alpha \mathbf{A}$. The entries of $\alpha \mathbf{A}$ are defined by:

$$(\alpha \mathbf{A})_{\mathbf{i,j}} = \alpha(\mathbf{A})_{\mathbf{i,j}}$$

More specifically this type of multiplication is called **left scalar multiplication**. It should be noted that the left scalar multiplication and the **right scalar multiplication**, $\mathbf{A}\alpha$, are equivalent on matrices, that is,

$$\alpha \mathbf{A} = \mathbf{A}\alpha$$

Example 2.7. *Suppose that we have a matrix $\mathbf{A}$ where $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \end{pmatrix}$ and we have some scalar $\alpha = 2$. Then $\alpha \mathbf{A} = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 6 & 10 \end{pmatrix}$*

Definition 2.15. The **transpose** of a matrix is an operation on matrices that switches the row and column indices of a matrix $\mathbf{A}$ by producing a new matrix that is denoted by $\mathbf{A^T}$. More formally, $a_{i,j}^T = a_{j,i}$.

It should be noted that if the dimension of a matrix $\mathbf{A}$ is $\dim(\mathbf{A}) = \mathbf{n} \times \mathbf{m}$, then $\dim(\mathbf{A^T}) = \mathbf{m} \times \mathbf{n}$.

Example 2.8. *Suppose that we have a matrix* $\mathbf{A}$ *where* $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, *then the transpose of A would be* $\mathbf{A^T} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

Definition 2.16. A square matrix whose transpose is equal to itself is called a **symmetric matrix**. That is, a matrix $\mathbf{A}$ is symmetric if

$$\mathbf{A^T} = \mathbf{A}$$

Definition 2.17. A square matrix whose transpose is equal to its negative is called a **skew-symmetric matrix**. That is, a matrix $\mathbf{A}$ is skew-symmetric if

$$\mathbf{A^T} = -\mathbf{A}$$

Example 2.9. *The matrix* $\boldsymbol{A} = \begin{pmatrix} 0 & 2 & -45 \\ -2 & 0 & -4 \\ 45 & 4 & 0 \end{pmatrix}$ *is a skew-symmetric matrix since* $\boldsymbol{A}^T = \begin{pmatrix} 0 & -2 & 45 \\ 2 & 0 & 4 \\ -45 & -4 & 0 \end{pmatrix} = -\boldsymbol{A}$

The **determinant** of a matrix is a value that is computed from the elements of a square matrix and contains information about the properties of the **linear transformation** that the matrix represents. Geometrically, it can be viewed as the scaling factor of volume represented by the linear

transformation. For example, the determinant of a 3×3 matrix is the volume of a parallelepiped that is spanned by the column or row vectors of the matrix.

The determinant of a matrix will be denoted in two ways. When using the matrix variable $\mathbf{A}$, the determinant is denoted $\det(\mathbf{A})$. In matrix form the determinant is denoted by using straight line brackets.

For 2×2 matrices the determinant is defined as:

$$det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{2}$$

Example 2.10. *The determinant of the matrix* $\mathbf{A} = \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix}$ *is found as follows:*

$$det(\mathbf{A}) = \begin{vmatrix} 2 & 4 \\ 3 & 7 \end{vmatrix} = \mathbf{2} \times \mathbf{7} - \mathbf{4} \times \mathbf{3} = \mathbf{2}$$

For 3×3 matrices the determinant is defined as:

$$det(\mathbf{A}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \tag{3}$$

$$= aei + bfg + cdh - ceg - bdi - afh. \tag{4}$$

In general, the determinant for any $n \times n$ matrix can be determined using the **Leibniz formula**:

$$det(\mathbf{A}) = \sum_{\sigma \in S_n} \left(sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right),$$

where S_n is the group of all permutations of the set $\{1, 2, 3, ..., n\}$, and $sgn(\sigma)$ is the signature of σ and has a value of $+1$ for even permutations

and -1 for odd permutations.

The following properties of the determinant are important to note for square matrices:

1. $\det(\mathbf{A}) \in \mathbb{R}$

2. $\det(\mathbf{I_n}) = 1$

3. $\det(\mathbf{A^T}) = \det(\mathbf{A})$

4. $\det(\mathbf{A}) \times \det(\mathbf{B}) = \det(\mathbf{AB})$, where $\mathbf{A}$ and $\mathbf{B}$ are the same size

2.3 Point set topology

Point set topology is a crucial component of the construction of structures such as Lie groups. It allows us to introduce notions of nearness, coordinates, and distances to what would otherwise be a set of values. For example, the set of real numbers in two dimensions, $\mathbb{R}^2$, is just a plane containing values. What gives the usefulness in everyday applications is the **topology** that is defined on it and the topological properties it has.

The tools developed in this subsection grant us the ability to describe the geometry of abstract spaces, enabling simpler computations. For example, a 3-dimensional space that is a topological space is the unit sphere. While it may be challenging to solve problems bounded to a sphere, a sphere is locally Euclidean to the real plane. As such, we can use homeomorphisms to map values to the real plane, do our calculations, and map back to the sphere. These processes and definitions are explored below.

Definition 2.18. Let X be a non-empty set. A **topology** on X is a collection of subsets $\tau \subseteq X$ which satisfy the following:

1. $\emptyset \in \tau$

2. $\tau \in X$

3. $\forall u, v \in \tau, (u \cap v) \in \tau$ (Closed under finite intersections)

4. $\forall C \subseteq \tau, \cup C \in \tau$ (Closed under arbitrary unions)

An element of τ is called an **open set**. It should be noted that we can define what an open set is in a variety of ways depending on how we construct the topology.

Defining a topology on a set is crucial as it is the weakest structure we can establish on a set to have the properties of convergence and continuity, or lack thereof.

Example 2.11. *Let $X = \{1, 2, 3\}$. Then an example of a topology on X would be $\tau = \{\emptyset, \{1\}, \{3\}, \{1, 3\}, \{1, 2, 3\}\}$*

Example 2.12. *The largest topology we can form on a set X is called the **discrete topology**. This topology is formed by using the **power set** of X, that is, the set of all subsets of X.*

Example 2.13. *One of the most important topologies we can define on a set is the **usual/standard topology** on the real numbers $\mathbb{R}^n$*
*The usual topology is formed by creating **open balls of dimension n** on $\mathbb{R}^n$. The open ball is an n-dimensional sphere of radius r around any arbitrary point $p \in \mathbb{R}^n$ and is constructed as follows.*

$$B_r(x) = \{y \in \mathbb{R}^n | \sqrt{\sum_{i=1}^{n}(y^i - x^i)^2} < r\} = \{y \in \mathbb{R} | d(x, y) < r\}$$

We can then define the topology as follows:

$$u \in \tau_{std} \text{ if and only if } \forall p \in u \; \exists r \in \mathbb{R}^+ \text{ such that } B_r(p) \subseteq u.$$

Proof. We want to show (w.t.s) that the standard topology is indeed a topology on $\mathbb{R}^n$. We need to check the following four properties:

1. $\emptyset \in \tau_{std}$ is vacuously true

2. W.t.s that $\mathbb{R}^n \in \tau_{std}$.

   By definition $\forall x \in \mathbb{R}^n \; \exists B_r(x) \subseteq \mathbb{R}^n$. Therefore $\mathbb{R}^n \in \tau_{std}$.

3. W.t.s that $\forall u, v \in \tau_{std}, \; u \cap v \in \tau_{std}$.

   Suppose $u, v \in \tau_{std}$

   Let $p \in u \cap v \rightarrow p \in u$ and $p \in v$

   Since $p \in u \; \exists r_1 \in \mathbb{R}^n$ s.t. $B_{r1}(p) \subseteq u$

   Since $p \in v \; \exists r_2 \in \mathbb{R}^n$ s.t. $B_{r2}(p) \subseteq v$

   Therefore $B_{min(r_1, r_2)}(p) \subseteq u$ and $B_{min(r_1, r_2)}(p) \subseteq v$

   Thus $B_{min(r_1, r_2)}(p) \subseteq u \cap v \implies u \cap v \in \tau_{std}$

4. W.t.s that $C \subseteq \tau_{std} \implies \cup C \in \tau_{std}$

   Let $u \in C \implies \forall p \in u \exists r \in \mathbb{R}^+$ s.t. $B_r(p) \subseteq u$

   Since $B_r(p) \subseteq u$ it directly follows that $B_r(p) \in \cup C$

   Thus $\cup C \in \tau_{std}$.

Therefore, the standard topology does indeed form a topology on $\mathbb{R}^n$ $\qquad \square$

Definition 2.19. A **topological space** is a non-empty set X paired with a topology.

Example 2.14. *Suppose $X = \{a, b, c\}$ and $T = \{\emptyset, \{a\}, \{a, b\}, X\}$. It can be easily shown that T is a topology on X by checking the above conditions. Thus, (X, T) is a topological space.*

Definition 2.20. A **base** or **basis** for a topology τ of a topological space (X, τ) is a set B of open subsets of X such that every element of τ is equal to a union of some subset of B. Informally, it is the minimum number of open sets required to create the topology. It should be noted that there can be multiple bases where some are bigger or smaller than others.

An alternate, but equally useful, way to construct a topology is by forming open neighborhoods.

Definition 2.21. Given a point p in a set X, we define a **neighborhood** as a subset S of X that includes a open set U that contains p. More formally,

$$p \in U \subseteq S$$

Definition 2.22. An **open neighborhood** is a neighborhood in which S is also an open set.

An important consequence of the neighborhood construction of open sets is that it will allow us to introduce the notion of limit points, which tell us information about the boundaries of a set. It should be noted that the open neighborhood construction is how we defined the standard topology on $\mathbb{R}^n$ above.

Example 2.15. *The n-dimensional set of real numbers, $\mathbb{R}^n$, forms a topological space with the standard topology, $(\mathbb{R}^n, \tau_{std,\mathbb{R}^n})$.*

An important property to note of topological spaces is that any subset S of a topological space (X, τ) is again a topological space with a new topology called the **subspace topology**. The subspace topology is defined as

$$\tau_S = \{S \cap U : U \in \tau\}$$

This property becomes very import when we introduce the notion of a topological space being **locally Euclidean**.

Definition 2.23. Let S be a subset of a topological space X. A point $p \in X$ is called a **limit point** of S if every neighborhood of p contains at least one point of S that is different from p.

The set of all limit points of a subset S is important because it contains all of the non-isolated points within S, called the **interior** of S, but also contains the points on the boundary of S.

Definition 2.24. The **closure** of a subset S of a set X is defined as the union of S with the set of all limit points of S.

Definition 2.25. If we have a topological space X, then a **cover** C of X is a collection of subsets U_i of X whose union is X. A **subcover** is a subset of C that still covers X. Additionally, we call C an **open cover** if each of its members is an open set, i.e., $U_i \in \tau$ where τ is the topology on X.

Definition 2.26. A **Hausdorff space** is a topological space X where for any two distinct points $p_1, p_2 \in X$ there exits neighborhoods of each that are disjoint.

Definition 2.27. A **second-countable space**, also called a **completely separable space**, is a topological space whose topology has a countable base. A **countable set** is a set with the same number of elements as some subset of the natural numbers.

Definition 2.28. A **homeomorphism** is a map f from one topological space, (X, τ_X), to another topological space, (Y, τ_Y), such that

1. f is a bijection

2. f is continuous

3. f has an inverse, f^{-1}, that is continuous

Homeomorphisms are crucial as when there is a homeomorphism between a topological space X and a topological space Y, then they have the same topological properties such as countability, separability, compactness, and connectedness. For example, if Y has a topological property that X does not have, then we know that no homeomorphism exists between the two spaces.

Definition 2.29. A **metrizable space** is a topological space that is homeomorphic to a metric space. More formally, a topological space (X, τ) is said to be metrizable if there is a metric $d : X \times X \to \mathbb{R}^+$ such that the topology induced by d is τ.

Definition 2.30. A topological space X is called **locally Euclidean** if there exists a positive integer n such that every point in X has a neighborhood that is homeomorphic to $\mathbb{R}^n$ equipped with the standard topology.

The topological properties and definitions above cover all of the topology pre-requisites to define manifolds, which are used in the construction of Lie groups.

2.4 Manifolds

Definition 2.31. A **topological manifold** X is a locally Euclidean Hausdorff space that is second-countable and metrizable . Some important properties to note are as follows:

1. The homeomorphism $\varphi : N \to V$, where N is a neighborhood in X and V is a neighborhood in $\mathbb{R}^n$, is called the **chart** or **coordinate system**.

2. The neighborhood N is the **domain** of the chart.

3. The image of N, $\varphi(N) \in \mathbb{R}^n$ is called the **coordinate** of the point p, where p is a point in the neighborhood N.

Example 2.16. *The set of real $n \times n$ matrices $M_n(\mathbb{R})$ forms a vector space that is isomorphic to $\mathbb{R}^{n^2}$ and contains an open subset $GL(n, \mathbb{R})$, called the general linear group. As such $M_n(\mathbb{R})$ is a topological manifold.*

Since topological manifolds are locally Euclidean, we introduce a new term to keep track of each chart that has neighborhood N_i and its corresponding homeomorphism φ_i. We define the **atlas** A of a topological manifold as the set containing all paired neighborhoods and homeomorphisms. That is,

$A = \{(N_i, \varphi_i) : i \in I\}$, where I is a collection of indices of the charts.

Definition 2.32. A **differentiable/smooth manifold** is a n-topological manifold with a smooth atlas. A smooth atlas is one in which every homeomorphism contained in the atlas is differentiable.

Example 2.17. *The surface given by $x^2 + y^2 + z^2 = 1$ is a differentiable manifold.*

2.5   Lie Groups and Lie Algebras

Definition 2.33. A **Lie group** is a group that is also a finite-dimensional differentiable manifold, in which the group operations of multiplication and

inversion are smooth maps. Smoothness of the group multiplication can be defined as follows:

$$(x, y) \mapsto x^{-1} y \tag{5}$$

Example 2.18. *An important example of a Lie group is the **general linear group***

$$\mathrm{GL}(n, \mathbf{R}) = \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \det A \neq 0 \right\} \tag{6}$$

Every linear algebraic group is a closed subgroup of the general linear group which is why it is a quintessential example of a Lie group.

Example 2.19. *The **orthogonal group** in dimension n, denoted by $O(n)$, is a subgroup of the general linear group that consists of distance-preserving transformations of Euclidean space of dimension n. The group operation is the composition of these transformations. The orthogonal group is defined on the reals $\mathbb{R}$ as*

$$O(n, \mathbb{R}) = \{ x \in GL(n, \mathbb{R}) | x^T x = x x^T = I \} \tag{7}$$

Example 2.20. *The **special orthogonal group** is a subgroup of the orthogonal group that consists of the transformations that are a rotation about θ*

$$\mathrm{SO}(2, \mathbf{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbf{R}/2\pi\mathbf{Z} \right\} \tag{8}$$

2.6 Lie Algebras and Lie Brackets

Definition 2.34. Let V, W, X be vector spaces. A **bilinear map** is a function $f : V \times W \to X$ such that for all $\alpha \in F$, $u, v \in V$, and $z, w \in W$,

$$f((\alpha u+v, w)) = \alpha f((u, w)) + f((v, w)) \text{ and } f((v, \alpha z+w)) = \alpha f((v, z)) + f((v, w)).$$

Observe that these two conditions are equivalent to saying that for each $v \in V$ and $w \in W$ the maps $T_v : W \to X$ defined by $T_v(z) = f((v, z))$ and $S_w : V \to X$ defined by $S_w(u) = f((u, w))$ are linear transformations. When $V = W$ we can check whether the bilinear map $f : V \times V \to X$ is one the following three important types:

1. **Symmetric**: for all $v, w \in V$, $f((v, w)) = f((w, v))$.

2. **Skew-symmetric**: for all $v, w \in V$, $f((v, w)) = -f((w, v))$.

3. **Alternating**: for all $v \in V$, $f((v, v)) = 0$.

Definition 2.35. A **Lie algebra** is a (finite-dimensional) vector space $\mathfrak{g}$ that is equipped with an alternating bilinear map $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying

$$[[u, v], w] + [[v, w], u] + [[w, u], v] = 0$$

for all $u, v, w \in \mathfrak{g}$. This equation is called the **Jacobi identity**, and such a map is called a **Lie bracket**.

Example 2.21. *Consider the zero-lie bracket defined on $\mathbb{R}^n$: $[u, v] = 0$ for all $u, v \in \mathbb{R}^n$. This makes $\mathbb{R}^n$ into a trivial Lie algebra. A Lie algebra is called **abelian Lie algebra** if its Lie bracket is identically 0.*

Example 2.22. *On $\mathbb{R}^3$ we define the Lie bracket to be the cross product.*
Let $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$. Define:

$$[\vec{u}, \vec{v}] = \vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).$$

3 Applications in Computing

3.1 Lie Group Integrators

Numerical methods are algorithms/procedures for finding numerical approximations of mathematical problems without symbolic computation. These methods are frequently used in performing computations that cannot be done symbolically or would be tedious to do so. As such, they are widely used throughout applied mathematics, engineering, and scientific fields.

A specific class of methods used frequently are numerical methods for ordinary differential equations, more informally called **numerical integration**.

Definition 3.1. An **ordinary differential equation** (ODE) is an equation involving a function y and its derivates. An ODE of order n is an equation of the form

$$F\left(x, y, y', ..., y^{(n)}\right) = 0, \tag{9}$$

where y is a function of x, y' is the first derivative of y with respect to x, and y^n is the nth derivative with respect to x [2].

ODEs are of incredible relevance to many scientific disciplines, including physics, biology, chemistry, and economics. This is a result of ODEs capturing the 'changes' that occur in these fields. Additionally, these disciplines are often concerned with using ODEs to make predictions about how a system will behave given different input conditions.

Without loss of generality, we can focus on first-order differential equations since higher-order ODEs can be converted into systems of first-order

differential equations by introducing additional variables. The focus of this section will be exploring first-order differential equations and how they can be improved by using symmetries.

Definition 3.2. A **first-order differential equation** is an ODE together with an initial condition which specifies the value of the unknown function y at a given point in the domain. A first-order differential equation has the form

$$y'(t) = f(t, y(t)), \ \ y(t_0) = y_0, \tag{10}$$

where f is a function $f : [t_0, \infty] \times \mathbb{R}^d \to \mathbb{R}^d$, and the initial condition $y_0 \in \mathbb{R}^d$.

Thus, numerical integration can be framed as a method that allows us to gain numerical approximations to first-order differential equations. A key point of numerical integration is that the results are approximations and prone to some level of error. Additionally, different numerical integration methods either prioritize computational speed or error. As such, choosing the proper integration method for a given problem can be challenging.

As seen in Figure 1, we can plot the analytical solution to a differential equation along with the approximations from different numerical integration methods to gain some intuition into how various integration methods provide different approximations. The primary factor that causes the differences illustrated in this Figure is the complexity of the integration methods used. The Euler method is the simplest method to utilize but has the most significant error, while the Runge-Kutta method requires additional complexity but produces far more accurate results [7]. Additionally, Figure 1 reveals how the Euler's method diverges rapidly while the Runge-Kutta

method remains relatively accurate during the same integration interval.
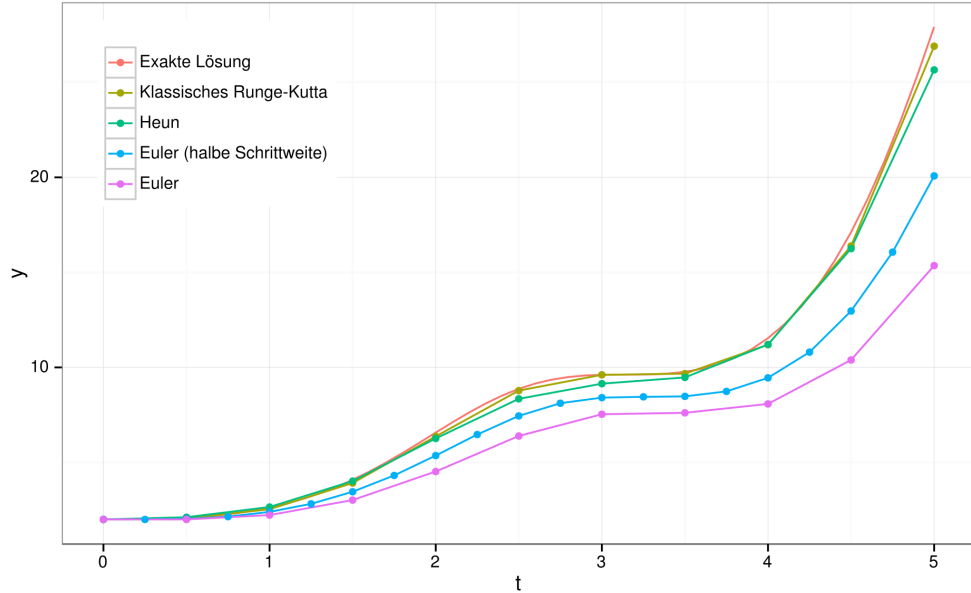

Figure 1: The output y, of the differential equation $\dot{y} = \sin(t)^2 y$ as predicted by different integrators contrasted against the exact solution

While Euler's method does accrue a significant error in relatively short intervals, it is one of the simplest integration methods to implement. As such, the Euler method can be used to demonstrate how improvements can be made to integration methods by using symmetries.

Definition 3.3. For a first-order differential equation of the form $\dot{y} = F(y)$ with the initial conditions $y(0) = y_0$ we can approximate the solution with **Euler's method** by taking a small time increment h, and approximating $y(h)$ using:

$$y_1 = y_0 + h\, F(y_0) \tag{11}$$

Example 3.1. *Suppose we have the differential equation $\dot{y} = x^2$ with the initial condition $y(0) = 1$ and we want to approximate the output of the ODE at $y(0.9)$. Euler's method of integration tells us that $y_1 = y(0) + h(0^2)$.*

Taking the step size $h = 0.3$, then

$y_1 = 1 + 0.3 \cdot 0^2 = 1$. *Similarly,*

$y_2 = y_1 + 0.3(x_1^2) = 1 + 0.3(0.3^2) = 1.027$, *and*

$y_3 = y_2 + 0.3(x_2^2) = 1.027 + 0.3(0.6^2) = 1.135$.

Thus $y(0.9) \cong 1.135$. We can now compare this to the analytical result of the differential equation. We can rewrite the ODE as $\frac{dy}{dx} = x^2$, thus by multiplying by dx we get $dy = x^2 dx$. Taking the integral of both sides results in $y + C_1 = \frac{x^3}{3} + C_2$, where C_1, C_2 are arbitrary constants. Now we can rewrite the equation as $y = \frac{x^3}{3} + C_3$, where $C_3 = C_2 - C_1$. Using the initial condition $y(0) = 1$ we find that $C_3 = 1$, thus the final result is $y(x) = \frac{x^3}{3} + 1$. Evaluating this equation at $y(0.9)$ yields $y(0.9) = \frac{0.9^3}{3} + 1 = 1.243$. This example shows that Euler's method was able to provide an approximate solution that was relatively close to the actual answer. More importantly, Euler's method would have worked and produced an approximate solution even if we could not have solved the equation analytically.

Another, but more complex way to think about Euler's method is to consider the constant vector field $F_{y0}(y) := F(y_0)$ which is obtained by parallel translating the vector $F(y_0)$ to all points of phase space [7]. As a consequence, it is necessary to compute the exact h-flow of this vector field starting at y_0. The phase space of a differential equation can be thought of as a fluid flow that shows you how the output of the equation will change based on different initial conditions. As such, the h-flow can be thought of as stepping along a specific streamline. An example of a phase space diagram more clearly illustrates this concept and is provided in Figure 2.

Definition 3.4. The **flow** on a set X is a group action of the additive group of real numbers on X. More formally, a flow is a map $f : X \times \mathbb{R} \to X$ such
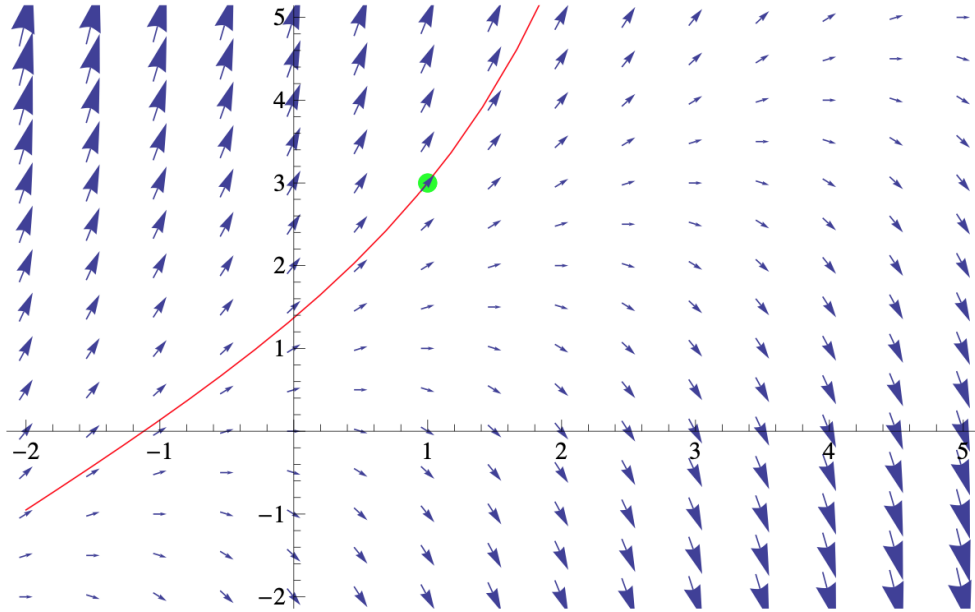
Figure 2: The phase space of the differential equation $y' = y - x$. The flow of the differential equation is represented with the blue vectors. Additionally the solution when the initial condition is $y(1) = 3$ is shown as a red curve.

that for all $x \in X$ and all real numbers y and z,

$$f(x, 0) = x \tag{12}$$

$$f(f(x, z), y) = f(x, y + z) \tag{13}$$

Shifting to this vector field based definition allows a clear transition to the modifications we can make to the Euler method using Lie groups. Lie group integration methods work using the same principle but benefit from allowing for more complex vector fields. As such, we can define a modified version of the Euler method called the **Lie-Euler method** using the following expression,

$$y_{n+1} = \exp(h F_{y_n}) y_n \tag{14}$$

28

where exp is the flow of a vector field.

Example 3.2. *Consider the following differential equation in $\mathbb{R}^3$ [6],*

$$\dot{\boldsymbol{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{pmatrix} = \begin{pmatrix} -y_2 + y_1 y_3^2 \\ y_1 + y_2 y_3^2 \\ -y3(y_1^2 + y_2^2) \end{pmatrix}$$

Note that $\frac{d}{dt}(y_1^2 + y_2^2 + y_3^2) = 0$. Thus the magnitude of $\boldsymbol{y} = (y_1, y_2, y_3)^T$ is time invariant in the exact solution. As such, the solution to the differential equation can be thought of as rotations on a sphere of radius $||\boldsymbol{y_0}||$. Thus, we can introduce the Lie group SO(3) to capture this symmetry in the integration. As such, we will need to formulate $\dot{\boldsymbol{y}}$ on the homogeneous space of a sphere. Now the Lie-Euler formulation for this problem becomes

$$\boldsymbol{y}_{n+1} = exp\left\{ h \begin{pmatrix} 0 & -(y_1^2 + y_2^2) & y_3(y_1 - y_2) \\ y_1^2 + y_2^2 & 0 & y_2(y_1 + y_2) \\ -y_3(y_1 - y_2) & -y_3(y_1 + y_2) & 0 \end{pmatrix} \right\} \boldsymbol{y_n}$$

Figure 3 shows the error for evaluating $\dot{\boldsymbol{y}}$ using the traditional Euler method and the Lie-Euler method [6]. Notably the error in the traditional Euler method quickly blows up, while Lie–Euler method stabilizes near the exact solution. While the Lie-Euler method did require us to identify the symmetry and recognize the suitable Lie group representation, it produced far more accurate results than the traditional Euler method.

As seen in the above example, by identifying the symmetries of a differential equation, we can use the corresponding Lie group representation to constrain the integration method. This allows for more accurate results that capture the essence of the original equation more accurately. This same principle can be applied to differential equations from various other disciplines if a symmetry is found.
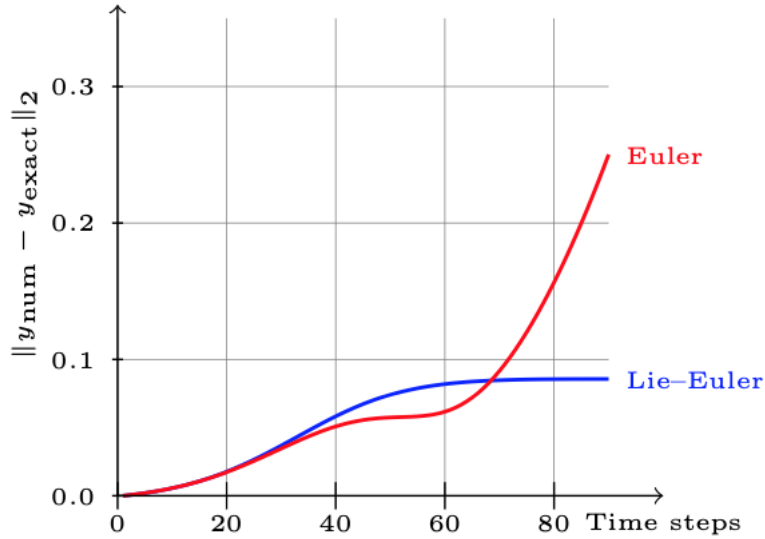
Figure 3: The error produced by traditional Euler method of integration compared to the error produced by the Lie-Euler method. Source: Berland

3.2 Energy Drift

The errors discussed in Section 3.1 that result from numerical methods have significant consequences for real-world systems. For example, computer simulations of mechanical systems suffer due to these gradual errors when performing long-term analyses. The result is a gradual change in the total energy of a closed system over time called **energy drift**. This is opposed to one of the fundamental laws of physics, which states that an isolated system's total energy should remain constant.

More specifically, the cause of this drift is error produced when analyzing the motion of the system. One of the key components of the total energy of the system is the kinetic energy, the energy associated with motion, which is governed by

$$E_{kinetic} = \frac{1}{2}m\vec{v}^{2} \tag{15}$$

where m is the mass of the system, and $\vec{v}$ is the velocity. Thus, when the

velocity has some small error $\delta\vec{v}$, the resulting energy of the system will be

$$E_{kinetic} = \sum m\vec{v}_{true}^2 + \sum m\delta\vec{v}^2, \tag{16}$$

where $\vec{v}_{true}$ is the actual velocity.

Due to this problem, there are significant benefits to choosing an integration method to reduce or eliminate the energy error. Integration methods are categorized into two groups, symplectic and non-symplectic. **Symplectic integrators** are integration methods designed to provide numerical solutions to Hamilton's equations, which are an alternate way of expressing equations of motion. The benefit of doing so is that symplectic integrators can take advantage of conserved quantities, i.e. symmetries, inherent to the formulation of Hamiltonians. As such, they significantly reduce the accrued error. The connection between conserved quantities and symmetries is explored in Chapter 4 of this paper.

While symplectic integrators are incredibly important, the necessary background information to fully explore them is outside the scope of this paper. The key takeaway of symplectic integrators is that they build on the symmetries of a system to provide more accurate solutions and are crucial in molecular dynamics, plasma physics, and celestial mechanics.

4 Applications in Physics

4.1 Noether's Theorem and Symmetry

Classical mechanics is a branch of physics concerned with describing motion. Fundamental to classical mechanics are the Lagrangian and Hamiltonian formalisms, which are an alternate way of expressing equations of motion. One of the essential facets of using these alternate formalisms comes from Noether's theorem, which we can use to connect continuous symmetries to conserved quantities. More specifically, Noether's theorem states that every differentiable symmetry of an action of a physical system has a corresponding conservation law. A conservation law is a measurable property of an isolated physical system that does not change as the system evolves over time. Common examples include the conservation of energy, conservation of momentum, and conservation of electric charge.

The goal of this section is to build up intuition as to how we can use Noether's theorem to identify the corresponding conservation law. As such, we first must define the necessary definitions from physics that are utilized when applying Noether's theorem.

Definition 4.1. The **Lagrangian** $\mathcal{L}$ of a system is defined as kinetic energy K of the system minus the potential energy U of the system. That is,

$$\mathcal{L} = K - U. \tag{17}$$

The kinetic energy is the energy it has by virtue of being in motion, while the potential energy of a system is the energy it has by virtue of its position relative to other bodies.

Definition 4.2. Generalized coordinates are a set of parameters that describe the configuration of a system with respect to some reference configuration. Notably, the minimum number of generalized coordinates needed to describe the system is given by the number of degrees of freedom.

These two definitions allow us to express Noether's theorem in a general manner that can be adapted to various physics problems. The critical component to recognize from these definitions is that we can describe the motion of a system using the energy and describe the location without specifying a coordinate system. Additional background information about Lagrangian mechanics can be found in classical mechanics textbooks.

Example 4.1. *Consider a pendulum of mass m hanging from a mass-less wire of length l from the ceiling. The mass experiences the acceleration to due gravity g and is free to rotate some angle θ. This scenario is depicted in Figure 4. In this system, there is one degree of freedom; thus, our generalized coordinate is θ. While we could specify a coordinate plane and express the motion in terms of its x and y coordinates, this leads to an unnecessary loss of generality. As such the kinetic energy of the system is,*
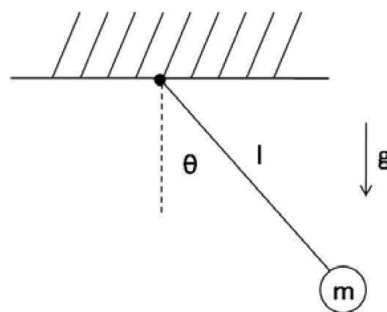

Figure 4: The error produced by traditional Euler method of integration compared to the error produced by the Lie-Euler method. Source: Berland

$$K = \frac{1}{2}m(v)^2 = \frac{1}{2}m(l\dot\theta)^2.$$

Similarly, we can express the potential energy U in terms of the generalized coordinate as

$$U = mgh = -mgl\cos(\theta).$$

These results follow from adapting the traditional expressions of kinetic and potential energy to depend on theta.

Thus, the Lagrangian of this system is

$$\mathcal{L} = K - U = \frac{1}{2}m(l\dot\theta)^2 + mgl\cos(\theta).$$

Now that we have established though an example of how Lagrangians and generalized coordinates can be applied to a specific problem, we can state Noether's theorem.

Definition 4.3. Assume we have a system described by generalized coordinates $q_1, ..., q_n$ and the Lagrangian $\mathcal{L}(q_1, ..., q_n, \dot{q}_1, ..., \dot{q}_n, t)$. Additionally assume that an infinitesimal transformation $q_i(t) \to q_i(t) + \delta q_i(t)$ is a symmetry transformation, that is, the action is invariant under this transformation. Then there is a conserved quantity of this system, J, such that

$$J = \sum_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i - F, \tag{18}$$

where $\dfrac{dF}{dt} = \Delta\mathcal{L}$ is the change in the Lagrangian under the transformation. This is called **Noether's theorem** in the Lagrangian formalism [9].

Example 4.2. *The Hydrogen Atom has a Coulomb potential given by the Hamiltonian $H = \frac{p^2}{2m} - \frac{k}{r}$, $k = e^2$. Since Noether's theorem is expressed using the Lagrangian formalism, the first step is to use a Legendre transform*

to derive the Lagrangian [9].

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2 + \frac{k}{|q|}, \ q = (x, y, z)^T$$

Analyzing the infinitesimal rotation ($\phi \to 0$) using the matrix representation of $SO(3)$ around the x axis:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \approx \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The approximation was performed by using a first-order Taylor expansion around $\phi = 0$. This allows us to express the transformation in the same form as is used in Noether's theorem:

$$x \to x + \delta x = x + 0 \tag{19}$$

$$y \to y + \delta y = y - z \tag{20}$$

$$z \to z + \delta z = z + y \tag{21}$$

Note that x is invariant under the transformation and that the Lagrangian itself is invariant under the transformation. Thus $\Delta\mathcal{L} = 0$. The conserved quantity is now expressed as

$$\frac{\partial\mathcal{L}}{\partial\dot{x}}\delta x + \frac{\partial\mathcal{L}}{\partial\dot{y}}\delta y + \frac{\partial\mathcal{L}}{\partial\dot{z}}\delta z = 0 - m\dot{y}z + m\dot{z}y = L_x$$

where L_x is the x component of angular momentum. Performing the analysis again using rotations about the y and z axis tells us that L_y and L_z are additionally conserved. Thus, the application of Noether's theorem to this problem revealed that the rotational symmetries that exist are a consequence

of the conservation of angular momentum.

We can see the connections between Noether's theorem and the symmetries that exist within a problem through this example. Additionally, it showed how analyzing a problem through the Lie group $SO(3)$ can reveal additional information from a relatively simple premise. These same principles hold for more complex examples. As such, Noether's theorem proves to be a practical application of Lie groups to higher-level physics such as quantum field theory.

4.2 The Lorentz Group

As seen in the above section, symmetries are a powerful tool in physics. Not only does it provide a compact way of expressing equations, but it can also be used to test the strength of the laws of physics. One of Lie Groups' most famous applications to physics is with the **Lorentz Group** [11].

The Lorentz group is the group containing all Lorentz transformations of Minkowski space-time. Minkowski space-time combines three-dimensional Euclidean space and time into a four-dimensional smooth manifold. The metric between any two events is independent of the inertial frame of reference. Only the relevant details of Minkowski space-time and Lorentz transformations will only be contained in this research. However, further information can be found in many undergraduate textbooks covering relativity [10].

4.2.1 Lorentz Transformation

The **Lorentz transformations** are a family of linear transformations from a coordinate frame in space-time to another frame that moves at a con-

stant velocity relative to the former and is the basis of special relativity. Essentially it describes how motion and time will be perceived in a different frame of reference. For example, if a spacecraft is traveling fast enough, the amount of time someone onboard the spacecraft experiences will be different from that of a stationary observer on Earth. This phenomenon is called **time dilation** and is described by the Lorentz transformations.

The Lorentz transformations are typically introduced as maps where,

$$t' = \gamma(t - \frac{vx}{c^2}) \tag{22}$$

$$x' = \gamma(x - vt) \tag{23}$$

$$y' = y \tag{24}$$

$$z' = z. \tag{25}$$

(t, x, y, z) and (t', x', y', z') are the coordinates of an event in two frames, where the prime frame is seen from the unprimed frame as moving with speed v along the x-axis, c is the speed of light, and $\gamma = (\sqrt{1 - \frac{v^2}{c^2}})$ is the Lorentz factor.

Example 4.3. *Suppose lightning strikes a tree and is observed by someone next to the tree at $t = 10$ microseconds. The lightning strike is also observed by someone in a rocket traveling in the positive x-direction with a velocity of $0.5c$. Relativity tells us that this lightning strike would not be observed at the same time for both observers. As such the Lorentz transform gives us the time that the rocket observes the lightning as*

$$t' = \frac{t - \dfrac{vx}{c^2}}{\sqrt{1 - \left(\dfrac{v}{c}\right)^2}} \qquad (26)$$

$$= \frac{(10 \times 10^{-6}s) - \dfrac{(0.5c)(0)}{c^2}}{\sqrt{1 - \left(\dfrac{0.5c}{c}\right)^2}} \qquad (27)$$

$$= 11.55 \times 10^{-6}s \qquad (28)$$

$$= 11.55 \; microseconds. \qquad (29)$$

This example illustrates the central notion of relativity, that time and space are relative.

While this form helps gain insights into how Lorentz transformations can be applied to common examples, physicists often prefer to analyze the effects of special relativity on space as a whole. It is valuable to form a group consisting of all Lorentz transforms where the group operation is composition. Notably, this forms a Lie group called the Lorentz group.

4.2.2 Forming the Lorentz Group

To simplify the derivation, we can choose the units of the speed of light to be $c = 1$ so that distance is measured in light-seconds and time is measured in seconds. This will make the rest of the Lorentz group's derivation easier to follow as it removes unit errors. Now, the first step in the

derivation is to express the event more succinctly using the four-vector $x^\mu = (t, x^i)$, where $\mu = 0, 1, 2, 3$.

Then, a Lorentz transformation on the four vector is expressed in Einstein index summation notation by

$$x^\mu \to L^\mu_\nu x^\nu \tag{30}$$

such that the transformation preserves the invariant interval $x^\mu x^\nu \zeta_{\mu\nu} = s^2$. $\zeta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and s is called the space time interval and is expressed by

$$s^2 = -t^2 + \vec{x}^{\,2}. \tag{31}$$

The space-time interval is a concept introduced to combine distances in space and time. The space-time interval's importance is that it remains the same in all reference frames, i.e., it is a symmetry.

Example 4.4. *The Lorentz boost is an example of a Lorentz transformation. The boost is defined as*

$$\begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}$$

where ϕ is the rapidity in the x direction. This transformation preserves $-t^2 + x^2$ and as such preserves $s^2 = -t^2 + x^2 + y^2 + z^2$. More specifically the matrix representation L of the boost transformation would be expressed as

$$\begin{pmatrix} \cosh\phi & \sinh\phi & 0 & 0 \\ \sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that the rapidity is defined in terms of the velocity v, as $v = \tanh\phi$. Thus we can rewrite the boost in the form

$$\begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} \gamma & \gamma\,v \\ \gamma\,v & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \tag{32}$$

where $\gamma = \frac{1}{\sqrt{1-v^2}} = \cosh\phi$. This form of the boost is the most similar to the form expressed at the beginning of Section 4.2.1.

The **Lorentz group** is the group containing all of the linear transformation matrices L^μ_ν that are Lorentz transforms. Notably the Lorentz group is a representation of the special orthogonal group $SO(3,1)$.

This group plays a crucial role in physics, especially at a higher level when quantum mechanics and relativity are brought together—as such, having a compact representation of time and space proves extraordinarily useful. Understanding its derivation and how it results from symmetries helps motivate higher levels of intuition when trying to understand these complex fields.

5 Applications in Control Theory

As the last example, we consider the use of symmetries and Lie groups in control theory. **Control theory** is the process by which engineers deal with the control of dynamical systems, a system whose state evolves with time over a state space according to some fixed rule [1].

A state space is the set of all possible configurations of a system and can be either discrete or continuous. An example of discrete but vast state space would be the set of all possible piece positions on a chessboard. A continuous example would be the position of a particle in some subset of R^3. The abstraction of state space is a valuable notion for control theory and is frequently used in artificial intelligence and game theory.

The objective of control theory is to develop a model governing how state inputs move the system to the desired state while minimizing error, delay, and control stability. The first step in doing so is to model the system using a State-Space representation. A State-Space representation is first-order differential equation that relates a mathematical model with input and output states [3].

Definition 5.1. A **dynamical system** is defined as a state space X, a set of times T, and a rule R that specifies how the state evolves with respect to time.

The rule is a function $R : X \times T \to X$, where $R = R(x,t)$, $x \in X$ is the initial state and $t \in T$ is some future time. Thus $R(x,t)$ tells us the state at time t given the initial state x.

If $R(x,t)$ obeys the superposition principle, the system is governed by linear differential equations and falls in the domain of **Linear Control**

Theory [4].

When $R(x, t)$ does not obey the superposition principle, the system is governed by nonlinear differential equations and falls in the domain of **Nonlinear control theory**. These types of problems are far more common as most real-world systems are not linear. Additionally, Lie groups can aid us in quickly determining if a non-linear system is controllable.

Definition 5.2. Controllability describes the ability of an input to move the internal state of a system from any initial state to any other final state in a finite time interval.

5.1   Linear and Nonlinear Control Theory

5.1.1   Linear

It can become cumbersome to represent complex systems using differential equations, especially if it has multiple inputs and outputs. The state-space representation of a system replaces an nth order differential equation with a single first-order matrix differential equation [8].

The state space representation for continuous linear systems is then given by the following two equations

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \tag{33}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) \tag{34}$$

where there are r inputs, and m outputs. Additionally, $\boldsymbol{x}$ is the state vector, a $n \times 1$ matrix,

$\boldsymbol{A}$ is a constant $n \times n$ matrix called the state matrix,

$\boldsymbol{B}$ is a constant $n \times r$ matrix called the input matrix,

$\boldsymbol{u}$ is $r \times 1$ vector called the input,

$\boldsymbol{C}$ is a constant $m \times n$ matrix called the output matrix,

$\boldsymbol{D}$ is a constant $m \times r$ matrix called the direct transition matrix,

$\boldsymbol{y}$ is a $m \times 1$ vector called the output.

Example 5.1. *Suppose we have a mass, m, on a frictionless surface attached to spring k_2 on the left side, and spring k_1 with a damper b on the right side. We apply some force $f_a(t)$ as the input and want to know the displacement z as the output. The aforementioned system is illustrated in figure (X). We can express the equations of motion as $m\ddot{x} + k_1 x + k_2 x - k_1 z = f_a$ and $b\dot{z} + k_1 z - k_1 x = 0$. There are three state equations resulting from the three energy storage elements m, k_1, and k_2. Since the energy of these three elements depends on x, $\dot{x}$, and z, the state variables are:*

$$q_1 = x \tag{35}$$

$$q_2 = \dot{x} \tag{36}$$

$$q_3 = z \tag{37}$$

From the equations of motion we can express the derivatives of the state variables as:

$$\dot{q}_1 = \dot{x} = q_2 \tag{38}$$

$$\dot{q}_2 = \ddot{x} = \frac{1}{m}(f_a - k_1 x - k_2 x + k_1 z) = \frac{1}{m}(f_a - k_1 q_1 - k_2 q_1 + k_1 q_3) \tag{39}$$

$$\dot{q}_3 = \dot{z} = \frac{k_1}{b}(x - z) = \frac{k_1}{b}(q_1 - q_3) \tag{40}$$

Thus the system can be expressed in matrix form as:

$$\dot{\boldsymbol{q}} = \boldsymbol{A}\boldsymbol{q} + \boldsymbol{B}u \tag{41}$$

$$y = \boldsymbol{C}\boldsymbol{q} + Du \tag{42}$$

where $u = f_a$ is the input, $y = z$ is the output and

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 \\ \dfrac{k_1 + k_2}{m} & 0 & \dfrac{k_1}{m} \\ \dfrac{k_1}{b} & 0 & \dfrac{-k_1}{b} \end{pmatrix}, \; \boldsymbol{B} = \begin{pmatrix} 0 \\ \dfrac{1}{m} \\ 0 \end{pmatrix}, \; \boldsymbol{C} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \; D = 0$$

As seen above, the motion equations are more compactly expressed and can easily be implemented into a computer system for solving.

5.1.2 Affine Non-Linear Systems

While linear control systems have many applications, most systems that occur in the physical world are not linear. Any model that does not fit the equation in the above section would be classified as nonlinear. An important family of nonlinear systems appear linear in the actions but nonlinear with respect to the state called **control-affine systems**

The formulation of control-affine systems can be achieved in terms of linear combinations of vector fields on a n dimensional differentiable manifold X. For these systems $\dot{x} = f(x, u)$ is of the form:

$$\dot{x} = h_0(x) + \sum_{i=1}^{m} h_i(x)\, u_i \tag{43}$$

where each h_i is a vector field on X and $u_i \in \mathbb{R} \in U \subseteq \mathbb{R}^m \; (m < n)$ is an

action variable that determines how much of $h_i(x)$ contributes to the result $\dot{x}$. Here, h_0 is called the **drift vector field** and h_i is called the **input vector field**.

It should be noted that in the case where X is a group, then by definition it is a Lie group. Thus, the drift vector field h_0 is an infinitesimal automorphism of X, and h_i for $i = 1...m$ are elements of the Lie algebra $L(X)$.

Definition 5.3. Let $X_1, ..., X_m$ denote smooth vector fields (v.f.) on a smooth n-dimensional manifold M. By definition, the **Lie algebra rank condition** at a point $p_0 \in M(LARC(p_0))$ is the property that $M_{p0} = SpanX(p_0) : X \in Lie(X_1, ..., X_m)$, where $Lie(X_1, ..., X_m)$ denotes the Lie algebra of v.f. generated by $X_1, ..., X_m$ [12].

Theorem 1. *Let $\Sigma(X, \mathcal{D})$ be a control-affine system*

1. *If Σ is controllable, then Σ satisfies the Lie algebra rank condition*

2. *if Σ satisfies the ad-rank condition, then Σ is controllable*

3. *If G is an Abelian Lie group, then Σ is controllable if Σ satisfies the rank condition. In particular, we can decide controllability with the rank condition for any Lie group*

$$X = T^n \times R^m, \ n \in \mathbb{N}, \ m \in \mathbb{N}, \tag{44}$$

where $T^n = S^1 \times S^1 ... S^1$, n times.

Example 5.2. *Let X be the Heisenberg group of dim 3*

$$X = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$$

with the Lie Algebra

$$L(X) = Span_{L.A} \left\{ Y^1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, Y^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, Y^3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}$$

As such, the Lie bracket is

$$[Y^1, Y^2] = Y^3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Considering the system with $\mathbb{D} = \{h_0 + uY^2 : u \in \mathbb{R}\}$ where the infinitesimal automorphism h_0 is defined by

$$h_0(x) = bY^3, \text{ for } x \in X$$

Then, the Semi direct product $< h_0 : \mathbb{H} >= Span_{LA}\{Y^2, Y^3\}$.

*Thus, Σ is **not controllable** [5].*

This example illustrates the power of Lie groups in the context of control theory. By analyzing the problem in terms of symmetries, we were able to gain insights into the Heisenberg group and its applicability to control theory. This same approach can be taken with other problems to significantly simplify what would often require a great deal of analysis.

6 Conclusion

The applications of Lie groups and general Lie theory are numerous and can produce powerful results in various fields including computing, physics, and control theory. These benefits are all a direct result of utilizing the symmetries that exist within each discipline.

For computing applications, these symmetries were used to adapt typical methods used for numerical integrations to ensure that solutions are within the problem's constraints. Physics applications resulted from Noether's theorem, which is used to understand where the conserved quantities, such as mass, energy, and momentum, arise. They are a direct by-product of the symmetries that exist in the universe. Finally, for control theory, we used symmetries to determine whether an essential class of non-linear control problems satisfies crucial properties that are of great importance to engineers.

Notably, in this research, many Lie group applications were excluded due to the high degree of field-specific information required to understand how Lie groups can be utilized. One such application that was omitted was the application to computer vision problems. Specifically, Lie groups can be used for camera motion estimation and the establishment of image geometry from said motion. For example, this technique can be used to create a 3D image via a drone's video feed. While outside of the scope of this paper, further research may illuminate this possible application of Lie groups, as well as many others. Although this paper is able to explain several pertinent applications that showcase the usefulness of Lie groups in several fields of study and industry, there are several applications that remain under-utilized and under-researched. This paper serves to encour-

age further inquiry into the utilization of Lie groups in a wide variety of relevant fields.

Index

References

[1] Math insight. URL `https://mathinsight.org/definition/dynamical_system`.

[2] Ordinary differential equation. URL `https://mathworld.wolfram.com/OrdinaryDifferentialEquation.html`.

[3] State-space representation, Apr 2021. URL `https://en.wikipedia.org/wiki/State-space_representation`.

[4] Control theory, Mar 2021. URL `https://en.wikipedia.org/wiki/Control_theory#Linear_and_nonlinear_control_theory`.

[5] V. Ayala and A. K. Hacibekiroglu. Examples of linear control systems on lie groups. *PRO MATHEMATICA*, X:19–20, 1996.

[6] H. Berland. Lie group and exponential integrators: Theory, implementation, and applications. Jul 2006. URL `https://core.ac.uk/download/pdf/52105963.pdf`.

[7] E. Celledoni, H. Marthinsen, and B. Owren. An introduction to lie group integrators – basics, new developments and applications. *Journal of Computational Physics*, 257:1040–1061, 2014. doi: 10.1016/j.jcp.2012.12.031.

[8] S. C. Erik Cheever. URL `https://lpsa.swarthmore.edu/Representations/SysRepSS.html`.

[9] G. Gentinetta. Continuous symmetries and lie algebras. *Algebra and Topology in Quantum Mechanics and Field Theory*, May

2020. URL `https://n.ethz.ch/~ggian/documents/2020_Gian_Gentinetta_Continuos_Symmetries_and_Lie_Algebras.pdf`.

[10] M. P. Hobson, G. Efstathiou, and A. N. Lasenby. *General relativity: an introduction for physicists*. Cambridge University Press, 2014.

[11] J. Huang. Lie groups and their applications to particle physics: A tutorial for undergraduate physics majors. Dec 2020. doi: arXiv: 2012.00834v1. URL `https://arxiv.org/pdf/2012.00834.pdf`.

[12] P. Morin and C. Samson. A characterization of the lie algebra rank condition by transverse periodic functions. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, volume 4, pages 3988–3993 vol.4, 2000. doi: 10.1109/CDC.2000.912338.

[13] J. J. Rotman. *An introduction to the theory of groups*. World Publishing Corporation, 2009.