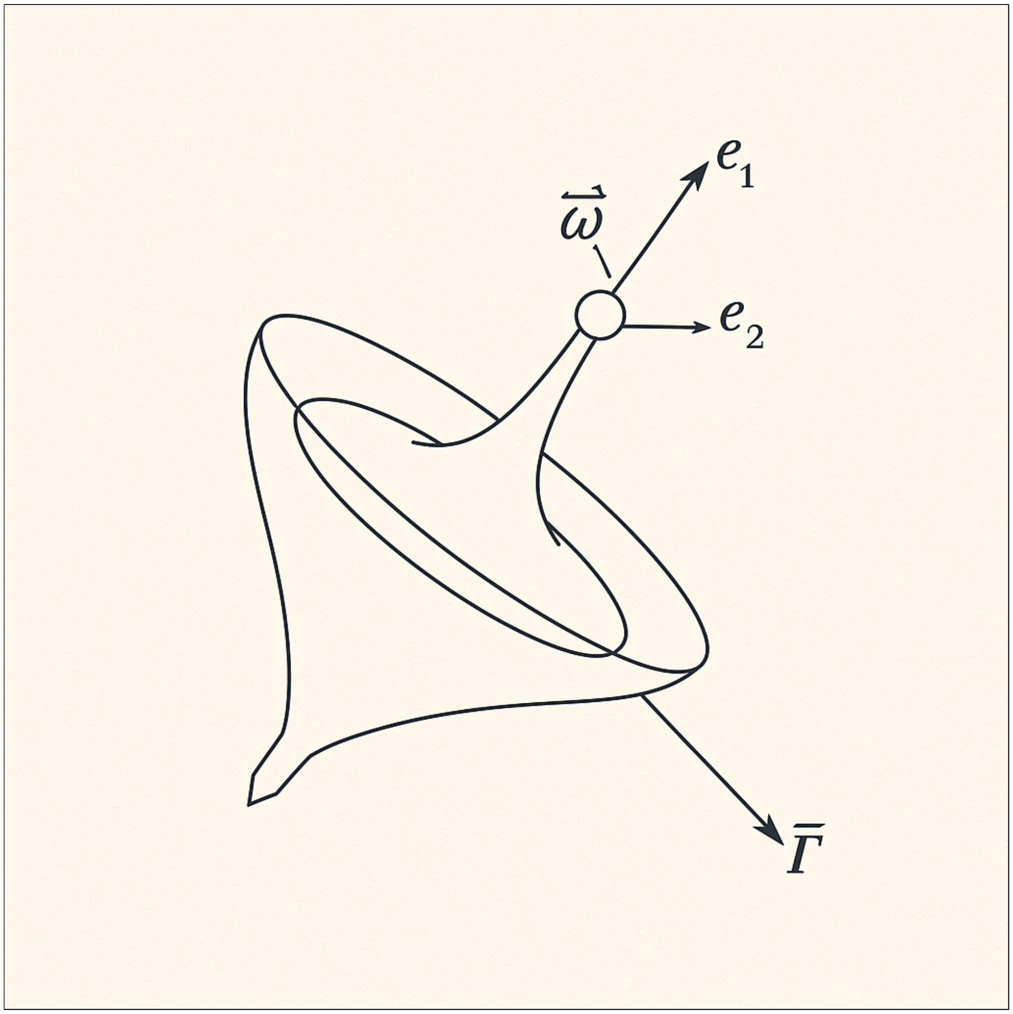
Lie-Poisson Reduction and the Heavy Symmetric Top

Parker Sherry

McGill University



May 25 2025

Acknowledgments

I'd like to extend a sincere thanks to my supervisor, William Holman-Bissegger, whose knowledge of geometry and topology was invaluable in the making of this paper. I'd also like to thank my DRP teammates, Mikey Baker and Amelie Chiasson-David, for supporting me and giving their feedback on my work. Finally, I'd like to thank my teacher, Prof. Katelin Schutz, for introducing me to this topic and passing on her love of geometric mechanics.

Many of the proofs in Sections 3 and 4 are taken from [1], which on the whole served as an extremely informative resource.

Contents

Introduction

Despite being one of the most familiar branches of physics, classical mechanics, in its modern formulation, can be quite difficult to deal with mathematically. Mathematicians have spent many years building upon the foundations laid by Newton, Lagrange, Hamilton, and Jacobi, and many subfields have sprung up as a result of their work. Take Floer theory, for example: an active field of research in modern mathematics, it was developed specifically to tackle problems in symplectic geometry, itself a byproduct of Hamilton's formulation of mechanics. Another, perhaps better-known, example comes from the work of Lorenz, Mandlebrot, and Feigenbaum in chaos theory, the study of unpredictable behaviour in deterministic, dynamical systems.

Speaking personally for a moment, this is one of the things which fascinates me so much about classical mechanics. The phenomena governed by the rules of mechanics are intuitive and commonplace: an apple falling due to gravity, a wheel rolling on an incline plane, the collision of two billiard balls on a smooth surface. And yet, peering deeper into the math, one encounters a level of complexity and sophistication rarely seen elsewhere. Though we can observe the apple falling from the tree, the system's roots run much deeper.

In this expository paper, I'm going to introduce some of modern mechanics' main players: manifolds and Lie groups. I'm going to define them, explain what geometric objects they can be equipped with, and talk about how to do mechanics on them. The danger, of course, is that this subject can become very abstract very fast. To remedy this, I've decided to stick to one concrete example which will help ground us on our journey: that of the heavy symmetric top. It's a great toy model to play with because it can be solved analytically, and it combines many of the tools one uses often when solving physics problems. My hope is that someone reading this, having gotten a sense that the material has become too heady, can come back to the image of the spinning top to get a feel for, say, how some quantity varies with time, or how infinitesimal nudges can accumulate to form continuous transformations.

The crown jewel of this paper, so to speak, is Lie-Poisson reduction, a method of simplifying the dynamics of a mechanical system by exploiting its geometric symmetries. As with any mathematical scenario involving a Legendre transform pair, we will start off with a simpler version, known as Euler-Poincaré reduction, which uses the Lagrangian framework of mechanics, before shifting gears to focus on the Hamiltonian framework.

Ultimately, my hope is that this paper is interesting to read for students of both mathematics and physics, and that there is something in it for everyone. Newcomers to mathematical mechanics have many, *many* opportunities to survey their surroundings only to find that the terrain has shifted underneath them, and that something which was once coarse and procedural has given way to something terrifying in its intricacy, awesome in its scope. I felt that way a lot when researching this topic. I hope you do, too.

Math and Physics Background

We assume that the reader is familiar with the following topics from mathematics and physics:

Math

Groups: an understanding of groups comprising the material of Algebra 1 (MATH 235 in McGill's Course Catalogue), although results like Lagrange's Theorem, Cayley's Theorem, etc. will not be used.

Vector Calculus: familiarity with calculus in $\mathbb{R}^n$, and with some of the more important function theorems (e.g. the Implicit Function Theorem).

Topology: the terms "open" and "closed" are almost always used in the context of $\mathbb{R}^n$. If a set U is open relative to a manifold M, we will specifically state that U is open relative to M.

Variational Calculus: Let $q : I \subset \mathbb{R} \to Q$ be a smooth path in Q. A **deformation** of $\mathbf{q}(t)$, denoted $\mathbf{q}(t, s)$, is a smooth map such that $s \in (-\varepsilon, \varepsilon)$, $\varepsilon > 0$, and $\mathbf{q}(t, 0) = \mathbf{q}(t)$ for all $t \in I$. The **variation** of the curve $\mathbf{q}(\cdot)$ corresponding to the deformation $\mathbf{q}(t, s)$ is given by

$$\delta \mathbf{q}(\cdot) = \frac{d}{ds}\bigg|_{s=0} \mathbf{q}(\cdot, s) \tag{1}$$

In addition, we will sometimes make use of the **variational derivative** in the context of the action functional, to be defined later. The variational derivative, denoted using "$\delta/\delta\mathbf{q}$" rather than "$d/d\mathbf{q}$" or "$\partial/\partial\mathbf{q}$", is defined in relation to the variation to a functional $\mathcal{S}$:

$$\delta \mathcal{S}[\mathbf{q}(\cdot)] = \langle \frac{\delta \mathcal{S}}{\delta \mathbf{q}}, \delta \mathbf{q} \rangle = \int_a^b \frac{\delta \mathcal{S}}{\delta \mathbf{q}} \cdot \delta \mathbf{q} \, dt \tag{2}$$

Physics

Newtonian mechanics: an understanding of the Newtonian formulation of dynamics. If the reader is unfamiliar with the finer details of Newtonian physics, such as the context in which quantities like linear momentum or energy are conserved, there should be no issue, but we will not elaborate on these finer points here.

Subscripts and Superscripts

Throughout this paper, subscripts and superscripts will be use to differentiate between covariant and contravariant quantities. Given a coordinate transformation, contravariant quantities (e.g. vectors) will change representation in a specific way dependent on the transformation, while covariant quantities (e.g. metric tensors) will change representation in a completely different way, also dependent on the transformation.

Contravariant quantities are represented using superscripts. Ex:

$$\mathbf{v} = (v^1, v^2, v^3)$$

Covariant quantities are represented using subscripts. Ex:

$$\mathbf{v} = (v_1, v_2, v_3)$$

1 Manifolds

"I peruse Manifold objects, no two alike, and every one good."

– **Walt Whitman, Song of Myself (1855)**

1.1 Preliminaries

It is good to start out with a heuristic approach to manifolds. Broadly speaking, manifolds are sets that locally resemble Euclidean space. We will begin this section by defining real submanifolds, subsets of $\mathbb{R}^n$ that satisfy the definition of a manifold. The important structures with which manifolds can be equipped are defined in this context, as it is easier to imagine things like tangent spaces and coordinate charts as living in an ambient space. Finally, to cover the topic more generally, we give the definition of an abstract manifold.

1.1.1 Submanifolds of $\mathbb{R}^n$

Submanifolds of $\mathbb{R}^n$ are defined via smooth embeddings or level sets. In the former case, given a function $\psi : A \subseteq \mathbb{R}^m \to \mathbb{R}^n$ whose Jacobian is injective at each point in its domain, a submanifold M may be related to ψ topologically using neighbourhoods of M. In the latter case, given a function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^k$ whose Jacobian is surjective at each point in its domain, the landscape of M can be inferred using the Implicit Function Theorem. To clarify these definitions, it will be important to introduce the following classes of functions.

Definition 1.1 (Submersions)**.** *A differentiable function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^k$ is a **submersion** at $\mathbf{x} \in A$ if the derivative $Df(\mathbf{x})$ is surjective. A function that satisfies this condition for all $\mathbf{x} \in A$ is called a submersion.*

Example 1.1. *Let $m \geq n$, with $\mathbb{R}^m = \mathbb{R}^n \times \mathbb{R}^{m-n}$. The linear projection map $\pi : \mathbb{R}^m \to \mathbb{R}^n$ given by $(x, y) \mapsto x$ is a submersion for all $(x, y) \in \mathbb{R}^m$. Indeed, for all $(x, y) \in \mathbb{R}^m$, we have*

$$D\pi(x, y) = \begin{pmatrix} \mathbb{I}_n & | & 0_{n\times(n-m)} \end{pmatrix} \tag{3}$$

Definition 1.2 (Immersion)**.** *A differentiable function $f : A \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is an **immersion** at $\mathbf{x} \in A$ if the derivative $Df(\mathbf{x})$ is injective. A function that satisfies this condition for all $\mathbf{x} \in A$ is called an immersion.*

Example 1.2. *Let $\psi : \mathbb{R}^2 \to S^2 \subset \mathbb{R}^3$ by given by*

$$\psi(\theta, \phi) = (\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)) \tag{4}$$

We have

$$D\psi(\theta, \phi) = \begin{pmatrix} -\sin(\phi)\sin(\theta) & \cos(\phi)\cos(\theta) \\ \sin(\phi)\cos(\theta) & \sin(\phi)\cos(\theta) \\ 0 & -\sin(\phi) \end{pmatrix} \tag{5}$$

For $\phi \neq k\pi$, $k \in \mathbb{Z}$, the two columns of $D\psi$ are linearly independent, so the restriction of ψ to the domain $\mathbb{R} \times (0, \pi)$ is an immersion. The restriction of ψ to the domain $(-\pi, \pi) \times (0, \pi)$ is an injective immersion, and defines the spherical coordinate system on S^2.

Definition 1.3 (Embedding)**.** *An **embedding** is an immersion $\psi : A \subseteq \mathbb{R}^m \to \mathbb{R}^n$ such that $\psi^{-1} : \psi(A) \to A$ is continuous.*

Example 1.3. *Let $\psi : \mathbb{R} \to \mathbb{R}^2$ by given by $\psi(\theta) = (\cos(\theta), \sin(\theta))$. We have $\psi'(\theta) = (-\sin(\theta), \cos(\theta)) \neq (0, 0)$ for any $\theta \in \mathbb{R}$, so ψ is an immersion. Its image is S^1.*

The restriction of ψ to $(0, 2\pi)$ is injective with continuous inverse $\psi^{-1} : (S^1 \backslash \{(1,0)\}) \to (0, 2\pi)$. Indeed, for any sequence $(\theta_k)_{k\in\mathbb{N}}$ in $(0, 2\pi)$ such that $\lim_{k\to\infty} \theta_k = \theta_0$, we have $\lim_{k\to\infty} \psi(\theta_i) = \psi(\theta_0)$ and $\lim_{k\to\infty} \psi'(\theta_i) = \psi'(\theta_0)$. Therefore ψ continuously differentiable on $(0, 2\pi)$, and so by the Inverse Function Theorem there exists $U(\theta_0) \subset (0, 2\pi)$ such that ψ^{-1} exists and is continuously differentiable on $U(\theta_0)$. Thus, $\psi^{-1}|_{(0,2\pi)}$ is continuous, and $\psi|_{(0,2\pi)}$ is an embedding.

Definition 1.4 (Diffeomorphism)**.** *A* ***diffeomorphism*** *is a differentiable map with a differentiable inverse.*

Remark that if f is a diffeomorphism, then $Df(\mathbf{x})$ is invertible for all $\mathbf{x}$. Thus, every diffeomorphism is also an immersion and a submersion. However, the converse does not always hold. The submersion given in Example 1.1, for instance, is not a diffeomorphism, since it has no well-defined inverse.

With these terms defined, we can formally introduce the notion of a submanifold of $\mathbb{R}^n$:

Theorem 1.1 (Equivalent Definitions of a Submanifold of $\mathbb{R}^n$)**.** *Let $M \subseteq \mathbb{R}^n$, $0 < m < n$, and $k = n - m$. The following are equivalent:*

1. *For every $\mathbf{a} \in M$ there exists a neighbourhood U of $\mathbf{a}$, a smooth submersion $f : U \to \mathbb{R}^k$, and $\mathbf{c} \in \mathbb{R}^k$ such that $M \cap U = f^{-1}(\mathbf{c})$*

2. *For every $\mathbf{a} \in M$ there exists a neighbourhood U of $\mathbf{a}$ such that $M \cap U$ is the graph of a smooth function expressing k of the standard coordinates in terms of the other m coordinates.*

3. *For every $\mathbf{a} \in M$ there exists a neighbourhood U of $\mathbf{a}$ and a smooth embedding $\psi : V \subseteq \mathbb{R}^m \to \mathbb{R}^n$ such that $\psi(V) = M \cap U$*

Proof. (1) $\implies$ (2) Take any $\mathbf{a} \in M$. There exists a neighbourhood U of $\mathbf{a}$, a smooth submersion f and $\mathbf{c} \in \mathbb{R}^k$ such that $M \cap U = f^{-1}(\mathbf{c})$. Note that

$$M \cap U = M \cap U \cap U = f^{-1}(\mathbf{c}) \cap U \tag{6}$$

If we let $f(\mathbf{a}) = \mathbf{c}$, then by the Implicit Function Theorem we have that $f^{-1}(\mathbf{c}) \cap U$ is the graph of a smooth function expressing k of the standard coordinates in terms of the other m coordinates. Since $f^{-1}(\mathbf{c}) \cap U = M \cap U$, we are done.

(2) $\implies$ (3) Again let $\mathbf{a} \in M$ and U a neighbourhood of $\mathbf{a}$. Suppose that $M \cap U$ is the graph of a smooth function g. Let $U \subset \mathbb{R}^n$ be written as $U_1 \times U_2$ for $U_1 \subset \mathbb{R}^m$ and $U_2 \subset \mathbb{R}^k$. Then we can write $M \cap U = \{(\mathbf{x}, g(\mathbf{x})) : \mathbf{x} \in U_1\}$. Now let $V = U_1$ and define $\psi : V \to \mathbb{R}^n$ by $\psi(\mathbf{x}) = (\mathbf{x}, g(\mathbf{x}))$. Then ψ is a smooth immersion and $\psi^{-1} : M \cap U \to V$ is continuous because it is equal to the projection $\pi : (\mathbf{x}, g(\mathbf{x})) \mapsto \mathbf{x}$. Therefore ψ is a smooth embedding.

(3) $\implies$ (1) Again let $\mathbf{a} \in M$ and U a neighbourhood of $\mathbf{a}$. Suppose we are guaranteed for each such $\mathbf{a}$ the existence of a smooth embedding $\psi : V \subset \mathbb{R}^m \to \mathbb{R}^n$ such that $\psi(V) = M \cap U$. Then $\dim(\psi(V)) = m$, so choose a basis of $\mathbb{R}^n$ such that, for all $\mathbf{x} \in M \cap U$, we have

$$\mathbf{x} = (x^1, \ldots, x^m, 0, \ldots, 0) \in U \tag{7}$$

Define

$$f : U \to \mathbb{R}^k, \quad (x^1, \ldots, x^n) \mapsto (x^{m+1}, \ldots, x^n) \tag{8}$$

It is trivial to show that f is smooth. It is also a submersion, as shown in Example 1.1. Finally, we note that

$$M \cap U = f^{-1}(\mathbf{0}) \tag{9}$$

$\square$

In each of the preceding cases we say that $M \subset \mathbb{R}^n$ is an m-dimensional manifold, denoted $\dim(M) = m$, with the codimension of M being $k = n - m$, denoted $\operatorname{codim}(M) = k$. Knowing that a manifold is m-dimensional is important as it allows one to construct local maps from neighbourhoods of M to subsets of $\mathbb{R}^m$. These local maps are called coordinate charts, and we define them here:

Definition 1.5 (Coordinate Chart)**.** *Let M be a submanifold of $\mathbb{R}^n$, let $\mathbf{a} \in M$ and $\psi : V \to M \cap U$ be the embedding corresponding to the neighbourhood U which contains $\mathbf{a}$. The inverse of ψ, denoted*

$$\varphi := \psi^{-1} : M \cap U \to V \tag{10}$$

is called a ***coordinate chart****, and the components of φ are called* ***local coordinates****.*

Remark*: By definition, every coordinate chart is continuous and has a continuous inverse.*

Remark*: Given two different coordinate charts φ_1 and φ_2 with overlapping domains, the map $\varphi_2 \circ \varphi_1^{-1}$ is called a* ***change-of-coordinates transformation****, or simply a change-of-coordinates.*

Example 1.4. *Consider $S^2 \subset \mathbb{R}^3$. Removing the point $(0, 0, 1)$, we can define the stereographic projection of $S^2 \backslash \{(0, 0, 1)\}$ onto $\mathbb{R}^2$ via*

$$(x, y, z) \mapsto \left(\frac{x}{1 - z}, \frac{y}{1 - z} \right) \tag{11}$$

This is a smooth coordinate chart on almost all of S^2 but is not well defined at the "North Pole". We can fix this by defining the denominator of the chart to be $1 + z$ instead, but we must then remove the point $(0, 0, -1)$. In fact, one can show that S^2 cannot be globally covered by any single chart which maps to $\mathbb{R}^2$.

Coordinate charts are useful because one can use them to perform computations in flat space, rather than on the manifold directly. An example of this is the evaluation of functions defined on the manifold:

Definition 1.6 (Representation in local coordinates)**.** *Let $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ be submanifolds with* $\dim(M) = q$, $\dim(N) = p$, *and $f : M \to N$. Given any two coordinate charts $\varphi_1 : M \cap U_1 \to V_1 \subset \mathbb{R}^q$ and $\varphi_2 : N \cap U_2 \to V_2 \subset \mathbb{R}^p$, the* ***representation of f in local coordinates*** *is given by*

$$\varphi_2 \circ f|_{M \cap U_1} \circ \varphi_1^{-1} \tag{12}$$

where it is assumed that $Im(f) \cap (N \cap U_2) \neq \emptyset$.

One can use the local representation of f to investigate its differentiability. In particular, with respect to the situation presented in Definition 1.6, we say that f is differentiable at $\mathbf{a} \in M$ if, for every φ_1 defined at $\mathbf{a}$ and every φ_2 defined at $f(\mathbf{a})$, the composition $\varphi_2 \circ f \circ \varphi_1^{-1}$ is differentiable at $\varphi_1(\mathbf{a})$. If this holds for all $\mathbf{a} \in M$, we say that f is differentiable. Note that, provided f^{-1} exists, one can use the local representation of f to check whether the function is a diffeomorphism.

1.1.2 Tangent and cotangent spaces

Since submanifolds of $\mathbb{R}^n$ locally resemble lower-dimensional Euclidean spaces, it is natural to ask how we can formally describe directions of motion or infinitesimal displacements along them. This leads us to the notion of the tangent space, which captures all possible velocities of curves that pass through a point on the manifold.

Definition 1.7 (Tangent vectors, tangent spaces)**.** *Let M be a submanifold of $\mathbb{R}^n$ and let $\mathbf{x} \in M$. A* ***tangent vector*** *to M at $\mathbf{x}$ is defined to be $g'(0)$ for some smooth path $g : \mathbb{R} \to M$ such that $g(0) = \mathbf{x}$. The* ***tangent space*** *of M at $\mathbf{x}$ is the set of all tangent vectors based at $\mathbf{x}$, denoted $T_{\mathbf{x}}M$*

$$T_{\mathbf{x}}M = \{(\mathbf{x}, \mathbf{v}),\ \mathbf{v} = g'(0),\ g : \mathbb{R} \to M,\ g(0) = \mathbf{x}\} \tag{13}$$

Before we move on, we prove an important result concerning the tangent spaces of embedded submanifolds. The corresponding result for level set submanifolds can be found in [1], though the reader is encouraged to try and prove the latter result for themselves.

Theorem 1.2 (Tangent spaces of parameterized manifolds)**.** *Let M be an m-dimensional submanifold of $\mathbb{R}^n$ and let $\mathbf{x} \in M$. If $M \cap U = Im(\psi)$ for some neighbourhood U of $\mathbf{x}$ and some embedding ψ with $\psi(\mathbf{a}) = \mathbf{x}$, then*

$$T_{\mathbf{x}}M = \{\mathbf{x}\} \times Im D\psi(\mathbf{a}) \tag{14}$$

Proof. Let $\psi : W \subset \mathbb{R}^m \to \mathbb{R}^n$ such that W open, $\mathrm{Im}\psi = M \cap U$ for $\mathbf{a} \in M$ and a neighbourhood U of $\mathbf{a}$. By Definition 1.7, it is clear that every tangent vector $\mathbf{v}$ at $\mathbf{x}$ is $g'(0)$ for some smooth $g : \mathbb{R} \to M \cap U$ with $g(0) = \mathbf{x}$. Define

$$\alpha(t) := \psi^{-1} \circ g(t) \tag{15}$$

Clearly $\alpha(0) = \mathbf{a}$ and $g = \psi \circ \alpha$. By the chain rule, we have $g'(0) = D\psi(\mathbf{a})\alpha'(0)$. Since α is an arbitrary path in W and W is open, $\alpha'(0)$ can take on any value in $\mathbb{R}^m$. Taking the union over all such vectors and pairing the resulting value of $g'(0)$ with $\mathbf{x}$, the result follows. □

The important thing to note about this result is that if $T_{\mathbf{x}}M$ is a tangent space, it is also a vector space. This means that linear combinations of tangent vectors yield other tangent vectors, and that tangent vectors can be scaled by real coefficients. This may seem obvious, but this result has been used in the scheme of larger proofs to great effect (see [2]).

Definition 1.8 (Tangent bundle, projection map)**.** *The **tangent bundle** of a submanifold M of $\mathbb{R}^n$, denoted TM, is the union of all the tangent spaces to M:*

$$TM = \bigsqcup_{\mathbf{x} \in M} T_{\mathbf{x}}M \tag{16}$$

*The **tangent bundle projection** is the map $\pi : TM \to M$ given by $(\mathbf{x}, \mathbf{v}) \mapsto \mathbf{x}$, i.e. the projection takes any tangent vector to its base point.*

Example 1.5. *Consider S^1 and its parametrization $\psi : (0, 2\pi) \to S^1$, $\psi(\theta) = (\cos(\theta), \sin(\theta))$. Let $a = \pi$ and $\mathbf{x} = \psi(a) = (-1, 0)$. We have*

$$D\psi(\pi) = \begin{bmatrix} -\sin(\pi) \\ \cos(\pi) \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \tag{17}$$

by the previous theorem, we have that

$$T_{\psi(\pi)}S^1 = \{(-1, 0)\} \times span\{(0, -1)\} = \{(-1, 0, 0, -\lambda) : \lambda \in \mathbb{R}\} \tag{18}$$

Generalizing this result to any $\mathbf{x} \in S^1$ and computing the tangent bundle, we find that

$$TS^1 = \{(x, y, -\lambda y, \lambda x) : x^2 + y^2 = 1,\ \lambda \in \mathbb{R}\} \tag{19}$$

Note that TS^1 is diffeomorphic to $S^1 \times \mathbb{R}$ via $(x, y, -\lambda y, \lambda x) \mapsto ((x, y), \lambda)$.

Example 1.6. *Consider S^2 and take any $\mathbf{a} \in S^2$. For any smooth path $g : I \subset \mathbb{R} \to S^2$ with $g(0) = \mathbf{a}$ we have that $g'(0)$ defines a tangent vector based at $\mathbf{a}$. It is easy to show that, for any such vector $\mathbf{v}$, we have $\mathbf{a} \cdot \mathbf{v} = 0$. Indeed, the analogous result for level sets of Theorem 1.1 says that the tangent space $T_{\mathbf{a}}S^2$ can be written as*

$$T_{\mathbf{a}}S^2 = \{(\mathbf{a}, \mathbf{v}) : \mathbf{a} \cdot \mathbf{v} = 0\} \tag{20}$$

The tangent bundle can then be written as

$$TS^2 = \{(\mathbf{a}, \mathbf{v}) \in \mathbb{R}^6 : ||\mathbf{a}|| = 1,\ \mathbf{a} \cdot \mathbf{v} = 0\} \tag{21}$$

In contrast to S^1, it is not possible to find a diffeomorphism from TS^2 to $S^2 \times \mathbb{R}^2$.

As with other vector spaces, every tangent space has a basis. To see this, consider the parametrization $\psi : W \subset \mathbb{R}^m \to \mathbb{R}^n$ of a smooth manifold M and suppose $\psi(\mathbf{q}) = \mathbf{x} \in M$. Since ψ is a parametrization, $D\psi(\mathbf{q})$ is injective. By the result of Theorem 1.2, $\mathrm{Im}D\psi = T_{\mathbf{x}}M$, and so $D\psi(\mathbf{q})$ is an isomorphism from $\mathbb{R}^m$ to $T_{\mathbf{x}}M$.

Definition 1.9 (Basis of a tangent space)**.** *In the above context, we define the following special tangent vectors at* $\mathbf{x}$*:*

$$\frac{\partial}{\partial q^i}(\mathbf{x}) := \frac{\partial}{\partial q^i}(\psi(\mathbf{q})) = D\psi(\mathbf{q})\mathbf{e}_i \tag{22}$$

where $\mathbf{e}_i$ *is the* i*th standard basis vector of* $\mathbb{R}^m$*. Note that* $\frac{\partial}{\partial q^i}(\mathbf{x})$ *is the* i*th column of* $D\psi(\mathbf{q})$*, so the vectors* $\{\frac{\partial}{\partial q^1}(\mathbf{x}), \ldots, \frac{\partial}{\partial q^m}(\mathbf{x})\}$ *form a basis for* $T_{\mathbf{x}}M$*.*

One can also track how tangent spaces change when one manifold is mapped onto another.

Definition 1.10 (Tangent map, tangent lift)**.** *Let* $M \subset \mathbb{R}^m$ *and* $N \subset \mathbb{R}^n$ *be submanifolds of* $\mathbb{R}^m$ *and* $\mathbb{R}^n$*, respectively. Let* $f : M \to N$ *be differentiable. The* ***tangent map*** *of* f *at* $\mathbf{x} \in M$ *is the map*

$$T_{\mathbf{x}}f : T_{\mathbf{x}}M \to T_{f(\mathbf{x})}N \tag{23}$$

$$(\mathbf{x}, \mathbf{v}) \mapsto \left(f(\mathbf{x}), \left.\frac{d}{dt}\right|_{t=0} f(g(t)) \right) \tag{24}$$

where $g(t)$ *is a path in* M *with* $g(0) = \mathbf{x}$ *and* $\frac{d}{dt}|_{t=0}(g(t)) = \mathbf{v}$*. The tangential component* $\mathbf{v}$ *of* $T_{\mathbf{x}}f$ *is often written* $df(\mathbf{x})$ *and is called the* ***derivative of*** f ***at*** $\mathbf{x}$*.*

Taken together, all of the maps $T_{\mathbf{x}}f$ *, for all* $\mathbf{x} \in M$*, define the* ***tangent lift*** *of* f*,*

$$Tf : TM \to TN \tag{25}$$

$$(\mathbf{x}, \mathbf{v}) \mapsto (f(\mathbf{x}), (T_{\mathbf{x}}f)(\mathbf{v})) \tag{26}$$

Proposition 1.1 (Path independence of tangent map)**.** *Let* M*,* N*,* f*, and* $\mathbf{x}$ *be defined as above. Then* $T_{\mathbf{x}}f$ *is independent of one's choice of path* $g : I \in \mathbb{R} \to M$*, so long as* g *satisfies* $g(0) = \mathbf{x}$ *and* $g'(0) = \mathbf{v}$*.*

Proof. Let $(T_{\mathbf{x}}f)(\mathbf{v})$ denote the action of $T_{\mathbf{x}}f$ on the second entry of the ordered pair $(\mathbf{x}, \mathbf{v})$. Working in ambient space coordinates (i.e. the Euclidean spaces in which M and N are embedded), define

$$F : U \subset \mathbb{R}^m \to \mathbb{R}^n,\ F|_M := f \tag{27}$$

An explicit computation gives

$$(T_{\mathbf{x}}f)(\mathbf{v}) = \left.\frac{d}{dt}\right|_{t=0} f(g(t)) = DF(\mathbf{x}) \cdot g'(0) = DF(\mathbf{x}) \cdot \mathbf{v} \tag{28}$$

□

Example 1.7. *Let* $F : \mathbb{R}^3 \to \mathbb{R}^3$ *be given in matrix representation by*

$$\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{29}$$

i.e. a rotation by an angle α *around the* z*-axis. Since* F *is linear,* $DF(\mathbf{x}) = F(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^3$*.*

Consider S^2*. It is easy to show that* F *maps every point on* S^2 *to another point on* S^2*, so let* $f := F|_{S^2}$ *and let* $\mathbf{x} = (1, 0, 0)$*. Then any tangent vector* $\mathbf{v}$ *based at* $\mathbf{x}$ *can be written as* $\mathbf{v} = (0, v^2, v^3)$*, and the action of the tangent map of* f *at* $\mathbf{x}$ *is given by*

$$T_{(1,0,0)}f : T_{(1,0,0)}S^2 \to T_{f((1,0,0))}S^2, \tag{30}$$

$$((1, 0, 0), (0, v^2, v^3)) \mapsto ((\cos\alpha, -\sin\alpha, 0), (-v^2 \sin\alpha, v^2 \cos\alpha, v^3)) \tag{31}$$

The preceding theorem and its associated example were carried out using coordinates written with respect to the standard basis of the ambient space. Now suppose we were to carry out computations using local coordinate charts. In particular, suppose we had a smooth embedding $\psi : W \subset \mathbb{R}^m \to \mathbb{R}^n$ such that W open and $\mathrm{Im}\psi = M \cap U$ for some neighbourhood U of M.

Lemma 1.1. *If $U \subseteq \mathbb{R}^n$ is open, then $TU = U \times \mathbb{R}^n$.*

Proof. Let $\mathbf{x} \in U$ and define

$$g : I \subset \mathbb{R} \to \mathbb{R}^n,\ g(t) = \mathbf{x} + t\mathbf{v},\ \mathbf{v} \in \mathbb{R}^n \tag{32}$$

Note that we require U open in this definition because $\dim(\partial U) < \dim(U)$. □

If $(\mathbf{q}^1, \dots, \mathbf{q}^m)$ are coordinates with respect to the standard basis of $\mathbb{R}^m$, then for any open set $W \subseteq \mathbb{R}^m$, we can write the coordinates of TW as $(\mathbf{q}^1, \dots, \mathbf{q}^m, \dot{\mathbf{q}}^1, \dots, \dot{\mathbf{q}}^m) \in W \times \mathbb{R}^m$. Since we have that $T_{\mathbf{x}}M = \mathrm{Im}D\psi(\mathbf{q})$ for $\psi(\mathbf{q}) = \mathbf{x}$, we can express every tangent vector based at $\mathbf{x}$ as $D\psi(\mathbf{q}) \cdot \dot{\mathbf{q}}$ for a unique $\dot{\mathbf{q}} \in \mathbb{R}^m$. Indeed,

$$D\psi(\mathbf{q}) \cdot \dot{\mathbf{q}} = D\psi(\mathbf{q}) \cdot \left(\sum_{i=1}^m \dot{q}^i \mathbf{e}_i \right) = \sum_{i=1}^m \dot{q}^i \left(D\psi(\mathbf{q}) \cdot \mathbf{e}_i \right) = \sum_{i=1}^m \dot{q}^i \frac{\partial}{\partial q^i}(\mathbf{x}) \tag{33}$$

The upshot of this expression is that one can also define tangent maps in terms of local coordinate charts $\varphi = \psi^{-1}$.

Definition 1.11 (Tangent map)**.** *Let $\psi_1 : W_1 \subset \mathbb{R}^p \to \mathbb{R}^m$, $\psi_2 : W_2 \subset \mathbb{R}^s \to \mathbb{R}^n$ be smooth embeddings such that $Im\psi_1 = M \cap U$, $Im\psi_2 = N \cap V$ for U, V open in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Let φ_1 be the coordinate chart corresponding to ψ_1 and define φ_2 analogously. For $\mathbf{x} \in M$ and $\mathbf{y} \in N$, write*

$$\mathbf{q} = \varphi_1(\mathbf{x}),\ \mathbf{r} = \varphi_2(\mathbf{y}) \tag{34}$$

Let $f : M \to N$ be differentiable. The tangent map of the representation of f in local coordinates is denoted Tf, and is given by

$$(\mathbf{r},\ \dot{\mathbf{r}}) = T(\varphi_2 \circ f \circ \varphi_1^{-1})(\mathbf{q},\ \dot{\mathbf{q}}) = (\varphi_2 \circ f \circ \varphi_1^{-1}(\mathbf{q}),\ D(\varphi_2 \circ f \circ \varphi_1^{-1})(\mathbf{q}) \cdot \dot{\mathbf{q}}) \tag{35}$$

Example 1.8. *Let $f : S^2 \to S^2$ be a rotation about the z-axis by an angle α as in Example 1.7. Recall the parametrization of S^2:*

$$\psi(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi) \tag{36}$$

In local coordinates, the action of f on S^2 is given by $f(\theta, \phi) = (\theta + \alpha, \phi)$. Computing the Jacobian of this transformation, we find

$$Df(\theta, \phi) = \begin{bmatrix} \frac{\partial}{\partial\theta}(\theta + \alpha) & \frac{\partial}{\partial\phi}(\theta + \alpha) \\ \frac{\partial}{\partial\theta}(\phi) & \frac{\partial}{\partial\phi}(\phi) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbb{I}_2 \tag{37}$$

Given this, we can define the tangent lift of f in local coordinates via

$$Tf(\theta, \phi, \dot{\theta}, \dot{\phi}) = (f(\theta, \phi),\ Df(\theta, \phi) \cdot (\dot{\theta}, \dot{\phi})) = (\theta + \alpha, \phi, \dot{\theta}, \dot{\phi}) \tag{38}$$

Remark that, given a transformation f that induces a change of coordinates from one set of local coordinates $\{\mathbf{q}\}$ to another set $\{\mathbf{r}\}$, one can write any component $\dot{r}^i$ of a tangent vector with respect to the original coordinates via a straightforward application of the chain rule:

$$\dot{r}^i = \sum_j \frac{\partial r^i}{\partial q^j} \dot{q}^j \tag{39}$$

Another important application of the chain rule has to do with scalar-output functions $f : M \to \mathbb{R}$ defined on a manifold M. Given the tangential component $\mathrm{d}f(\mathbf{x})$ of a tangent map $T_{\mathbf{x}}f$, one has, for any $\mathbf{v} \in T_{\mathbf{x}}M$,

$$\mathrm{d}f(\mathbf{x}) \cdot \mathbf{v} = (f \circ g)'(0) \in \mathbb{R} \tag{40}$$

Thus, for any function f which assigns real numbers to points on a manifold, we have that the tangential component of the map $T_{\mathbf{x}}f$ also acts as a real-valued map, this time mapping the tangent space of a point on the manifold to the real numbers. It is also linear, meaning that it forms part of the space which is *dual* to $T_{\mathbf{x}}M$. Recall that the dual space of any real vector space V, denoted V^*, is the set of linear maps from V to $\mathbb{R}$. This set is also a vector space, and its elements are called covectors.

Definition 1.12 (Cotangent Vectors, cotangent Spaces)**.** *The* ***cotangent space*** *to M at $\mathbf{x}$ is $T^*_{\mathbf{x}}M := (T_{\mathbf{x}}M)^*$, the dual space of $T_{\mathbf{x}}M$. The* ***cotangent bundle*** *T^*M of M is defined as*

$$T^*M = \bigsqcup_{\mathbf{x}\in M} T^*_{\mathbf{x}}M \tag{41}$$

The cotangent bundle projection is the map $\pi : T^*M \to M$ given by $(\mathbf{x}, \mathbf{p}) \mapsto \mathbf{x}$.

The interaction between cotangent vectors and functions is defined via a canonical bilinear form called the **natural pairing**. Given a vector space V, for any $\mathbf{v} \in V$ and $\alpha \in V^*$, we define the natural pairing of v and α as the function $\langle \cdot, \cdot \rangle : V^* \times V \to \mathbb{R}$ given by

$$\langle \alpha, \mathbf{v} \rangle = \sum_i \alpha_i v^i \tag{42}$$

In this sense, one can construct maps between cotangent spaces which are "dual" to their corresponding tangent maps.

Definition 1.13 (Cotangent map, cotangent lift)**.** *Let $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ be submanifolds of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Let $f : M \to N$ be differentiable. The* ***cotangent map*** *of f at $\mathbf{x} \in M$ is the map*

$$T^*_{\mathbf{x}}f : T^*_{f(\mathbf{x})}N \to T^*_{\mathbf{x}}M \tag{43}$$

which satisfies, for all $\mathbf{v} \in T_{\mathbf{x}}M$ and $\alpha \in T^*_{f(\mathbf{x})}N$,

$$\langle (T^*_{\mathbf{x}}f)(\alpha), \mathbf{v} \rangle = \langle \alpha, (T_{\mathbf{x}}f)(\mathbf{v}) \rangle \tag{44}$$

If f is a diffeomorphism, we also define the ***cotangent lift*** *of f at $\mathbf{x}$ to be the cotangent map of f^{-1} at $f(\mathbf{x})$, which is the map*

$$T^*_{f(\mathbf{x})}(f^{-1}) : T^*_{\mathbf{x}}M \to T^*_{f(\mathbf{x})}N \tag{45}$$

given by

$$\langle T^*_{f(\mathbf{x})}(f^{-1})(\alpha), \mathbf{v} \rangle = \langle \alpha, T_{f(\mathbf{x})}(f^{-1})(\mathbf{v}) \rangle \tag{46}$$

Remark that if f maps M to N, its cotangent map maps "backwards" from $T^*_{f(\mathbf{x})}N$ to $T^*_{\mathbf{x}}M$, whereas the cotangent lift maps "forwards" from $T^*_{\mathbf{x}}M$ to $T^*_{f(\mathbf{x})}N$. For this reason, when f is a diffeomorphism, it is more common to use the cotangent lift than the cotangent map.

1.2 Vector Fields

We have seen how to equip a manifold M with tangent spaces and their corresponding cotangent spaces. These spaces comprise a colossal number of vectors: more advantageous in many situations would be a specific arrangement of vectors, assigned smoothly to each point $\mathbf{x} \in M$. To this end, we introduce the concept of a vector field. While some readers will have seen vector fields in the context of $\mathbb{R}^n$, these objects live in full generality on abstract manifolds. As such, we will suppress the phrase "an m-dimensional submanifold of $\mathbb{R}^n$" going forward, with the intention that readers continue to think of these objects as being embedded in an ambient, Euclidean space.

Definition 1.14 (Vector field)**.** *A* ***vector field*** *on a manifold M is a map $X : M \to TM$ such that $X(\mathbf{z}) \in T_{\mathbf{z}}M$ for all $\mathbf{z} \in M$. Vector fields can be added together and scaled by functions $k : M \to \mathbb{R}$, satisfying*

$$(X_1 + X_2)(\mathbf{z}) = X_1(\mathbf{z}) + X_2(\mathbf{z}), \quad (kX)(\mathbf{z}) = k(\mathbf{z})X(\mathbf{z}) \tag{47}$$

Definition 1.15 (Integral curve)**.** *Let X be a vector field. An* ***integral curve*** *of X is a differentiable map $c : I \subset \mathbb{R} \to M$ such that I open and such that $c'(t) = X(c(t))$ for all $t \in I$*

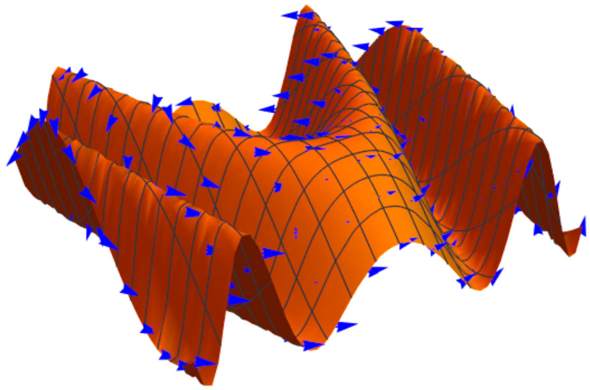


Figure 1: A submanifold of $\mathbb{R}^3$, parametrized by the function $\psi(x, y) = \sin(x + y^2)$, equipped with the vector field $X(x, y) = (-y, x, 0)$.

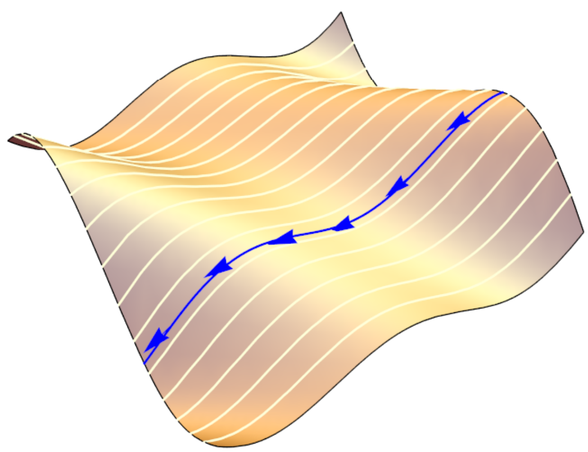


Figure 2: A submanifold of $\mathbb{R}^3$ equipped with a vector field. The integral curve along the set of blue vectors corresponds to a particular solution of the differential equation implied by the vector field.

Note that implicit in the definition of an integral curve is an ordinary differential equation. This is because integral curves represent particular solutions to ODEs. We will see many examples in the following chapters of functions defined on manifolds which induce a vector field, the solutions of which can be thought of as integral curves corresponding to that vector field.

Example 1.9. *Let X be the vector field on $\mathbb{R}^2$ defined by $X(x, y) = (-y, x) \in T_{(x,y)}\mathbb{R}^2$. Defining a path $c : I \subset \mathbb{R} \to \mathbb{R}^2$ such that $c(t) = (x(t), y(t))$, we have that*

$$\begin{cases} \dot{x} = -y, \\ \dot{y} = x \end{cases} \tag{48}$$

This set of differential equations has a family of solutions of the form $c(t) = (A\cos(t + \omega), A\sin(t + \omega))$

As highlighted in the previous example, it is often helpful to imagine "families of integral curves," i.e. time evolution along several integral curves at once. To this end we introduce the notion of a flow:

Definition 1.16 (Time-t flow)**.** *Let X be a differentiable vector field on a manifold M. A **flow** of X is a differentiable map $\Phi : U \times I \to M$, where $I \subseteq \mathbb{R}$, $t_0 \in I$, and U is a subset of M such that, for any $z \in U$, the map $\Phi(z,t)$ is an integral curve of X with $\Phi(z,t_0) = z$.*

*The **time-t flow** of X is the "fixed time" map $\Phi_t(z) := \Phi(z,t)$.*

Theorem 1.3 (Properties of the time-t flow)**.** *Let Φ be a flow of a differentiable vector field X on a manifold M. Then:*

1. *$\Phi_{t_0} = Id$ for all $z \in U$.*
2. *$\Phi_{t+s} = \Phi_t \circ \Phi_s$ for all $t, s \in I$.*
3. *Φ is a diffeomorphism for all $t \in I$.*

We will use interchangeably the phrases "a flow of X" and "the flow of X", as it is almost always clear what the "natural" flow of a vector field is in context.

We can also generalize the notion of a tangent map, or tangent lift, to any vector field defined on a manifold. These operations are called a push-forward or pull-back, respectively, and they operate on all vectors in a vector field simultaneously.

Definition 1.17 (Push-forward, pull-back)**.** *Let M, N be manifolds and let $\varphi : M \to N$ be a diffeomorphism. The **push-forward** of the vector field X on M by φ is the vector field $\varphi_* X$ on N, defined by*

$$\varphi_* X = T\varphi \circ X \circ \varphi^{-1} \tag{49}$$

In local coordinates, with $\mathbf{r} = \varphi(\mathbf{q})$, the push-forward is written as

$$(\varphi_* X)(\mathbf{r}) = D\varphi(\mathbf{q}) \cdot X(\mathbf{q}) = \frac{d\mathbf{r}}{d\mathbf{q}} \cdot X(\mathbf{q}) \tag{50}$$

*The **pull-back** of a vector field Y on N by φ is the vector field $\varphi^* Y$ on M defined by*

$$\varphi^* Y = T\varphi^{-1} \circ Y \circ \varphi \tag{51}$$

An application of the Inverse Function Theorem gives, in local coordinates,

$$(\varphi^* Y)(\mathbf{q}) = (D\varphi(\mathbf{q}))^{-1} \cdot Y(\mathbf{r}) = \frac{d\mathbf{q}}{d\mathbf{r}} \cdot Y(\mathbf{r}) \tag{52}$$

More pertinently, we may also want to measure the rate of change of one vector field X "along" another vector field Y. We can do this using the pull-back of a flow in the following way:

Definition 1.18 (Lie derivative)**.** *Let X, Y be differentiable vector fields on a manifold M, with Φ the flow of X. The **Lie derivative** of Y along X is given by*

$$\mathcal{L}_X Y = \left.\frac{d}{dt}\right|_{t=0} \Phi_t^* Y \tag{53}$$

Note that if f is a smooth scalar field on M, we can also compute the Lie derivative of f along X without loss of generality via

$$(\mathcal{L}_X f)(z) = df(z) \cdot X(z) \tag{54}$$

Example 1.10. *Consider the vector fields on $\mathbb{R}^2$ given by*

$$X(x,y) = (x,y), \quad Y(x,y) = (1,0) \tag{55}$$

The first vector field corresponds to a system of ODEs solved by $(x(t), y(t)) = (xe^t, ye^t)$*. Taking* $\Phi_t(x, y) = (xe^t, ye^t)$*, we have that*

$$D\Phi_t(x, y) = \begin{bmatrix} e^t & 0 \\ 0 & e^t \end{bmatrix} = e^t \cdot \mathbb{I}_2 \tag{56}$$

Computing the Lie derivative of Y *along* X*, we find*

$$\begin{aligned} \mathcal{L}_X Y &= \frac{d}{dt}\bigg|_{t=0} (\Phi_t^*(x, y)Y) && (57) \\ &= \frac{d}{dt}\bigg|_{t=0} (D\Phi_t(x, y))^{-1} \cdot Y(\Phi_t(x, y)) && (58) \\ &= \frac{d}{dt}\bigg|_{t=0} \begin{bmatrix} e^-t & 0 \\ 0 & e^-t \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} && (59) \\ &= \begin{bmatrix} -1 \\ 0 \end{bmatrix} && (60) \end{aligned}$$

1.3 A Note on Abstract Manifolds

So far we have described manifolds as being embedded in an ambient Euclidean space. This description suffices for the purposes of this paper, but it is worth noting that the definitions we have introduced can be weakened to divorce the concept of a manifold from any notion of embedding. Here is a topological definition of a manifold, taken from [3]:

Definition 1.19 (Topological manifold)**.** *A topological space* M *is locally Euclidean of dimension* n *if every point* p *in* M *has a neighborhood* U *such that there is a homeomorphism* ϕ *from* U *onto an open subset of* $\mathbb{R}^n$*. We call the pair* (U, ϕ) *a chart,* U *a coordinate neighborhood or a coordinate open set, and* ϕ *a coordinate map or a coordinate system on* U*. We say that a chart* (U, ϕ) *is centered at* $p \in U$ *if* $\phi(p) = 0$*.*

A topological manifold is a Hausdorff, second countable, locally Euclidean space. It is said to be of dimension n *if it is locally Euclidean of dimension* n*.*

There are benefits characterizing manifolds this way. For example, the topological definition of an abstract manifold makes it easier to quickly identify which topological properties a manifold can and cannot have. For example, the set defined by two intersecting lines in $\mathbb{R}^2$ is not a manifold, since the set is 1-dimensional and would be homeomorphic to a subset of $\mathbb{R}$. However, removing the point of intersection would split the set into four connected components, whereas its image under homeomorphism would only have two connected components. Since the number of connected components of a set is preserved under homeomorphism, the set cannot be a manifold.

2 Lagrangian and Hamiltonian Mechanics

"Go, wondrous creature! mount where science guides;
Go, measure earth, weigh air, and state the tides;
Instruct the planets in what orbs to run;
Correct old time, and regulate the sun."

- Alexander Pope, An Essay on Man (1733)

In Newtonian mechanics, given a point particle m whose position $\mathbf{x}(t)$ and velocity $\dot{\mathbf{x}}(t)$ are defined at $t = t_0$, the position of the particle at any later time $t \geq t_0$ can be completely determined by solving the equation

$$\ddot{\mathbf{x}}(t) = \frac{1}{m}\sum_i \mathbf{F}_i = -\frac{1}{m}\sum_i \nabla V_i \tag{61}$$

While in principle this equation represents three ordinary differential equations which can be solved using standard methods, in practice this process is often complicated by the inclusion of constraints or dissipative forces. Worse still, it is often the case that physicists do not want to explicitly solve for the equation of motion of the particle, but want to gleam some *aspect* of its motion. An engineer, for example, might want to calculate the minimum angular velocity at which a spinning top, rotating about its pivot point, remains standing. If forced to solve for the top's motion in its entirety, huge amounts of computational time and energy are wasted.

In Lagrangian and Hamiltonian mechanics, a dynamical system is characterized by its energy, rather than the forces acting on it. This seems a small change, but the ideas underpinning it are subtle, and the equations which arise from these frameworks shepherd in new, coordinate-independent methods of studying mechanics.

2.1 Lagrangian Mechanics

Lagrangian mechanics uses a function $L : TQ \to \mathbb{R}$, known as a *Lagrangian*, to generate equations of motion for a system parametrized by time. Here, Q is a manifold, often called the configuration space of the system, and TQ is the tangent bundle of Q. According to Hamilton's Principle, the path taken by a system through its configuration space has a *stationary action*. We define the action functional, $\mathcal{S}$, here, along with the Euler-Lagrange equations, the equations of motion implied by $\delta\mathcal{S} = 0$.

Definition 2.1 (Action functional)**.** *Let $q : \mathbb{R} \to Q$ be a parametrized path in Q and let $L : TQ \to \mathbb{R}$ be a Lagrangian on Q. Consider the set $\mathcal{Q}[t]$ of all parametrized paths in Q. The **action functional** $\mathcal{S} : \mathcal{Q}[t] \to \mathbb{R}$ is defined as*

$$\mathcal{S}[\mathbf{q}(t)] = \int_a^b L(\mathbf{q}(t), \dot{\mathbf{q}}(t))dt \tag{62}$$

We would like to act upon $\mathcal{S}$ in such a way as to obtain the same equations of motion we would by carrying out a Newtonian analysis of $\mathbf{q}(t)$, a trajectory on Q. To this end, we prove the following Lemma:

Lemma 2.1. *Let $(\mathbf{q}, \dot{\mathbf{q}}) \in TQ$, L, and $\mathcal{S}$ be defined as above. Then,*

$$\delta\mathcal{S} = 0 \iff \int_a^b \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\mathbf{q}\, dt = \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\Big|_a^b \tag{63}$$

Proof. Recall that the variation $\delta\mathcal{S}$ is equivalent to evaluating the derivative of the deformation of $\mathcal{S}$ at $s = 0$. Noting that $\delta\mathcal{S}[\mathbf{q}(t)] = d/ds|_{s=0}\mathcal{S}[\mathbf{q}(t, s)]$, $d/dt\, \delta\mathbf{q} = \delta\dot{\mathbf{q}}$, and that it is possible to differentiate under the integral sign with respect to s, we have

$$\delta\mathcal{S} = \frac{d}{ds}\bigg|_{s=0} \int_a^b L(\mathbf{q}(t,s), \dot{\mathbf{q}}(t,s))dt \tag{64}$$

$$= \int_a^b \left[\frac{\partial L}{\partial \mathbf{q}} \cdot \delta\mathbf{q} + \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\dot{\mathbf{q}}\right] dt \tag{65}$$

$$= \int_a^b \left[\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)\right] \cdot \delta\mathbf{q}\, dt + \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\bigg|_a^b \tag{66}$$

If $\delta\mathcal{S} = 0$, we obtain the desired result. □

The integrand in Equation 66, when set equal to zero, comprises a vectorized version of the Euler-Lagrange equations,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \qquad i = 1, \ldots, n \tag{67}$$

for $\dim(Q) = n$. We prove an important Lemma about the Euler-Lagrange equations here.

Lemma 2.2. *Let Q be an m-dimensional submanifold of $\mathbb{R}^n$ and let $q(t)$ be a parametrized path on Q. Then every Newtonian system*

$$\ddot{\mathbf{q}} = -\frac{1}{m}\sum_j \nabla V_j \tag{68}$$

is equivalent to the Euler Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \qquad i = 1, \ldots, m \tag{69}$$

For the Lagrangian $L : TQ :\to \mathbb{R}$ defined by

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}m||\dot{\mathbf{q}}||^2 - \sum_j V_j(\mathbf{q}) \tag{70}$$

Proof. For L as defined above, we have

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \frac{d}{dt}(m\dot{q}^i) + \sum_j \frac{\partial V_j}{\partial q^i} = m\ddot{q}^i + \sum_j \frac{\partial V_j}{\partial q^i} = 0 \tag{71}$$

Dividing through by m, we obtain the ith component of $\ddot{\mathbf{q}}$ and $1/m\sum_j \nabla V_j$. □

Given the results of the preceding Lemma, we give the interpretation of $\delta\mathcal{S}$ as employed in the Lagrangian formulation of mechanics as follows: given a path $\mathbf{q}(t)$ which represents the configuration of a system as it evolves in time between two specified states $\mathbf{q}(t = a)$ and $\mathbf{q}(t = b)$, we say that $\mathbf{q}(t)$ is the true evolution of the system (that is, the evolution which is equivalent to the one derived by Newtonian analysis) if it extremizes the action functional. Since $\mathbf{q}(a)$ and $\mathbf{q}(b)$ are fixed, we have $\delta\mathbf{q}(a) = \delta\mathbf{q}(b) = 0$, and $\delta\mathcal{S}$ as given in Lemma 2.1 becomes

$$\delta\mathcal{S} = \int_a^b \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\mathbf{q}\, dt = 0 \tag{72}$$

Which implies

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} = 0 \tag{73}$$

for aribtrary variations $\delta\mathbf{q}$. This is the mathematical formulation of Hamilton's Principle.

To complete the Lagrangian formulation of mechanics on manifolds, we must show that arbitary paths on manifolds satisfy Hamilton's Principle irrespective of our choice of local coordinates.

Theorem 2.1 (Euler-Lagrange equations)**.** *Let Q be a manifold. For any smooth Lagrangian $L : TQ \to \mathbb{R}$, a path $\mathbf{q}(t)$ in Q satisfies Hamilton's Principle if and only if it satisfies the Euler-Lagrange equations in every local coordinate system.*

Proof. Let $\{t_0 = a, t_1, \ldots, t_r = b\}$ be a partition of $[a, b]$ such that each subpath from t_{i-1} to t_i is contained in the domain of a single coordinate chart.

$\Longrightarrow$ Sketch: assume $\mathbf{q}(t)$ satisfies Hamilton's Principle. Then

$$\left.\frac{\partial}{\partial s}\right|_{s=0} \int_a^b L(\mathbf{q}(t,s), \dot{\mathbf{q}}(t,s))dt = 0 \tag{74}$$

For any variation $\delta\mathbf{q}(t)$ which vanishes at a and b, we have

$$\int_a^b \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\mathbf{q}\, dt = 0 \tag{75}$$

The issue is that, for an arbitrary coordinate chart φ_i with domain U_i, we are not guaranteed to have $\mathbf{q}(a)$, $\mathbf{q}(b) \in U_i$, so the "local" endpoints, $\mathbf{q}(t_{i-1})$ and $\mathbf{q}(t_i)$, may not be fixed. However, given any such subset U_i, any variation of the ith subpath which *does* have $\delta\mathbf{q}(t_{i-1}) = \delta\mathbf{q}(t_i) = 0$ can be extended to a variation of the entire path that is trivial for $t \notin [t_{i-1}, t_i]$. This extension is not necessarily smooth, but we can approximate it arbitrarily closely by smooth variations $\delta\tilde{\mathbf{q}}(t)$:

$$\int_{t_{i-1}}^{t_i} \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\mathbf{q}\, dt \approx \int_a^b \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\tilde{\mathbf{q}}\, dt = 0 \tag{76}$$

With a little analytical work, one can show that the left hand side is exactly zero, and so the Euler-Lagrange equations are satisfied in the local coordinate patch containing $\{\mathbf{q}(t) : t \in [t_{i-1}, t_i]\}$.

$\Longleftarrow$ Suppose that $\mathbf{q}(t)$ satisfies the Euler–Lagrange equations everywhere. Then for each $[t_{i-1}, t_i]$, we have that $\mathbf{q}(t)$ satisfies

$$\int_{t_{i-1}}^{t_i} \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right) \cdot \delta\mathbf{q}\, dt = \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{t_{i-1}}^{t_i} \tag{77}$$

Let $\delta\mathbf{q}$ be a variation to the path such that $\delta\mathbf{q}(a) = \delta\mathbf{q}(b) = 0$. Then, the restriction of $\delta\mathbf{q}$ to any subinterval $[t_{i-1}, t_i]$ is a variation of the ith subpath, so the equation above holds for every subpath. Adding the contribution of each integral, one obtains the full action functional, and adding the corresponding right-hand terms gives

$$\sum_{i=0}^{r} \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{t_{i-1}}^{t_i} = \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{t_0=a}^{t_1} + \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{t_1}^{t_2} + \cdots + \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{t_{r-1}}^{t_r=b} = \left.\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q}\right|_{a}^{b} = 0 \tag{78}$$

□

2.1.1 The Lagrangian Vector Field

Consider the Euler-Lagrange equations. Expanding the time derivative gives the equivalent equations

$$\frac{\partial^2 L}{\partial \dot{\mathbf{q}}^2}\ddot{\mathbf{q}} + \frac{\partial^2 L}{\partial \dot{\mathbf{q}} \partial \mathbf{q}}\dot{\mathbf{q}} - \frac{\partial L}{\partial \mathbf{q}} = 0 \tag{79}$$

For sufficiently "nice" Lagrangians, we can invert this equation to obtain an expression for $\ddot{\mathbf{q}}$. In this spirit, we say that a Lagrangian L is **regular** if

$$\det \frac{\partial^2 L}{\partial \dot{\mathbf{q}}^2} \neq 0 \tag{80}$$

for all $(\mathbf{q}, \dot{\mathbf{q}}) \in TQ$. Once satisfied, this condition nets us the following equation:

$$\ddot{\mathbf{q}} = \left(\frac{\partial^2 L}{\partial \dot{\mathbf{q}}^2}\right)^{-1} \left(\frac{\partial L}{\partial \mathbf{q}} - \frac{\partial^2 L}{\partial \dot{\mathbf{q}} \partial \mathbf{q}} \dot{\mathbf{q}}\right) \tag{81}$$

This implies a system of ODEs of the form

$$\frac{d}{dt}\mathbf{q} = \dot{\mathbf{q}} \tag{82}$$

$$\frac{d}{dt}\dot{\mathbf{q}} = \left(\frac{\partial^2 L}{\partial \dot{\mathbf{q}}^2}\right)^{-1} \left(\frac{\partial L}{\partial \mathbf{q}} - \frac{\partial^2 L}{\partial \dot{\mathbf{q}} \partial \mathbf{q}} \dot{\mathbf{q}}\right) \tag{83}$$

We have already shown that the Euler-Lagrange equations are local coordinate-independent. Therefore, the system of equations above are also coordinate-independent, and so form a vector field on TQ.

Definition 2.2 (The Lagrangian vector field Z_L)**.** *The **Lagrangian vector field** Z_L on TQ is defined, in local coordinates, by the system of equations*

$$\frac{d}{dt}\mathbf{q} = \dot{\mathbf{q}} \tag{84}$$

$$\frac{d}{dt}\dot{\mathbf{q}} = \left(\frac{\partial^2 L}{\partial \dot{\mathbf{q}}^2}\right)^{-1} \left(\frac{\partial L}{\partial \mathbf{q}} - \frac{\partial^2 L}{\partial \dot{\mathbf{q}} \partial \mathbf{q}} \dot{\mathbf{q}}\right) \tag{85}$$

Therefore, an equivalent statement of Theorem 2.1 goes as follows: if L is regular, then a path $\mathbf{q}(t)$ in Q satisfies the Euler-Lagrange equations if and only if $(\mathbf{q}(t), \dot{\mathbf{q}}(t))$ is a solution of the Lagrangian vector field on TQ.

2.2 Hamiltonian Mechanics and Poisson Manifolds

An important, complementary view of mechanics was given by Hamilton, 54 years after Lagrange's work was published. Hamiltonian mechanics extends Lagrangian mechanics by considering functions defined on the co-tangent space of a system, rather than on its tangent space. Working with respect to this space allows one to exploit some of its special geometric properties; in particular, it allows one to come up with novel ways of describing the time evolution of a mechanical system.

We must first define the way in which we normally translate from a Lagrangian viewpoint to a Hamiltonian one. With this in mind, we give here the definition of the Legendre transform:

Definition 2.3 (Legendre transform)**.** *Let (Q, L) define a mechanical system with configuration space Q and smooth Lagrangian L. The quantity*

$$\mathbf{p}(\mathbf{q}, \dot{\mathbf{q}}, t) := \frac{\partial L}{\partial \dot{\mathbf{q}}} \tag{86}$$

*is called the canonical momentum, or simply the momentum. For a given $t \in \mathbb{R}$, the **Legendre transform** (or Legendre transformation) of L is defined as the smooth map $\mathbb{F}L : TQ \to T^*Q$ given by*

$$\langle \mathbb{F}L(\mathbf{q}, \dot{\mathbf{q}}_1), (\mathbf{q}, \dot{\mathbf{q}}_2) \rangle := \left(\mathbf{q}, \left.\frac{d}{ds}\right|_{s=0} L(\mathbf{q}, \dot{\mathbf{q}}_1 + s\dot{\mathbf{q}}_2) \right) = \left(\mathbf{q}, \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}_1) \cdot \dot{\mathbf{q}}_2 \right) \tag{87}$$

We are interested in Legendre transformations which are invertible, i.e. which allow us to solve for $\dot{\mathbf{q}}$ as a function of $\mathbf{p}$ and vice versa. To this end, we define the following sub-class of Lagrangians:

Definition 2.4 (Hyperregular)**.** *A Lagrangian L is **hyperregular** if $\mathbb{F}L$ is a diffeomorphism. Remark that hyperregularity implies regularity.*

We will assume for the rest of this chapter that all Lagrangians are hyperregular. This allows us to define the corresponding Hamiltonian functions:

Definition 2.5 (Energy function)**.** *The* ***energy function*** *for a Lagrangian* $L : TQ \to \mathbb{R}$ *is* $E : TQ \to \mathbb{R}$ *defined by*

$$E(v) = \langle \mathbb{F}L(v), v \rangle - L \tag{88}$$

In local coordinates,

$$E(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}) \tag{89}$$

Definition 2.6 (Hamiltonian)**.** *The* ***Hamiltonian*** *corresponding to* L *is* $H : T^*Q \to \mathbb{R}$ *defined by*

$$H := E \circ (\mathbb{F}L)^{-1} \tag{90}$$

In local coordinates, H *is given by*

$$H(\mathbf{q}, \mathbf{p}) = \mathbf{p} \cdot \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p}) - L(\mathbf{q}, \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})) \tag{91}$$

Theorem 2.2. *For any hyperregular Lagrangian* $L : TQ \to \mathbb{R}$ *let* H *be the corresponding Hamiltonian. The Euler–Lagrange equations, in any tangent-lifted local coordinates* $(\mathbf{q}, \dot{\mathbf{q}})$ *on* TQ*, are equivalent to Hamilton's equations of motion*

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \tag{92}$$

in the corresponding cotangent-lifted coordinates $(\mathbf{q}, \mathbf{p})$ *on* T^*Q*.*

Proof. We first express Hamilton's equations of motion with respect to the underlying Lagrangian,

$$\frac{\partial H}{\partial \mathbf{q}} = \mathbf{p} \cdot \frac{\partial \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} - \frac{\partial L}{\partial \mathbf{q}} - \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \frac{\partial \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} = -\frac{\partial L}{\partial \mathbf{q}} \tag{93}$$

$$\frac{\partial H}{\partial \mathbf{p}} = \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p}) + \mathbf{p} \cdot \frac{\partial \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})}{\partial \mathbf{p}} - \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \frac{\partial \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} = \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p}) \tag{94}$$

where in both cases we have used the definition of $\mathbf{p}$ to cancel terms. The first of Hamilton's equations of motion is obtained explicitly through the expansion above. For the second, we note that the Legendre-transformed Euler Lagrange equations read

$$\frac{d}{dt}(\mathbf{p}) - \frac{\partial L}{\partial \mathbf{q}} = 0 \tag{95}$$

Which gives us the second of Hamilton's equations. □

Corollary 2.1. *Hamilton's equations of motion define a vector field* X_H *on* T^*Q*, called the* ***Hamiltonian vector field****. This vector field can be thought of as the push-forward of the Lagrangian vector field by the Legendre transform, i.e.* $H = (\mathbb{F}L)_* Z_L$*.*

2.2.1 Poisson Brackets, Poisson Manifolds

Upon first learning about the Hamiltonian formulation of mechanics, one might object that Hamilton's equations of motion can hardly be said to represent an improvement upon the results one can get by Lagrangian mechanics alone. As previously stated, there would be very little need for Hamiltonian mechanics in many cases were it not for the geometric properties of the spaces on which Hamiltonians are defined. These spaces, when equipped with an operator called a Poisson bracket, are referred to as Poisson manifolds, and we will presently introduce them in detail.

Definition 2.7 (Poisson Bracket)**.** *A* ***Poisson bracket*** *on a manifold* P *is a bilinear, skew-symmetric operator* $\{\cdot, \cdot\} : C^\infty(P, \mathbb{R}) \times C^\infty(P, \mathbb{R}) \to C^\infty(P, \mathbb{R})$ *satisfying*

1. $\{\{F, G\}, H\} + \{\{H, F\}, G\} + \{\{G, H\}, F\} = 0$ *(Jacobi identity)*

2. $\{FG, H\} = F\{G, H\} + \{F, H\}G$ *(Leibniz identity)*

Example 2.1. *Let* $(\mathbf{q}, \mathbf{p}) = (q^1, \ldots, q^n, p_1, \ldots, p_n) \in T^*Q$*, and let* $F, G : T^*Q \to \mathbb{R}$*. The* ***canonical Poisson bracket*** *on* Q *is given by*

$$\{F, G\} = \sum_i \left(\frac{\partial F}{\partial q^i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q^i} \right) \tag{96}$$

So long as n *is not too big, one can easily check that the canonical Poisson bracket satisfies both of the conditions in Definition 2.7*

The following theorem highlights the importance of Poisson brackets in Hamiltonian analysis:

Theorem 2.3. *Let* $\{\cdot, \cdot\}$ *be the canonical Poisson bracket on* T^*Q*, and let* $F : T^*Q \to \mathbb{R}$ *be differentiable. Hamilton's equations for a given* $H : T^*Q \to \mathbb{R}$ *are equivalent to*

$$\dot{F} = \{F, H\} \tag{97}$$

along all integral curves. More precisely, $(\mathbf{q}(t), \mathbf{p}(t))$ *is a solution of Hamilton's equations if and only if*

$$\frac{d}{dt}(F(\mathbf{q}(t), \mathbf{p}(t))) = \{F, H\}(\mathbf{q}(t), \mathbf{p}(t)) \tag{98}$$

Proof. We have

$$\frac{dF}{dt} = \sum_i \left(\frac{\partial F}{\partial q^i} \dot{q}^i + \frac{\partial F}{\partial p_i} \dot{p}_i \right) \tag{99}$$

$$= \sum_i \left(\frac{\partial F}{\partial q^i} \frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial H}{\partial q^i} \right) \tag{100}$$

$$= \{F, H\} \tag{101}$$

□

Note that $(\mathbf{q}(t), \mathbf{p}(t))$ above is also a solution to the Hamiltonian vector field, so the time derivative of F can be alternatively written as

$$\dot{F}(\mathbf{q}(t), \mathbf{p}(t)) = \langle \mathrm{d}F(\mathbf{q}(t), \mathbf{p}(t)), \frac{d}{dt}(\mathbf{q}(t), \mathbf{p}(t)) \rangle = \mathcal{L}_{X_H} F(\mathbf{q}(t), \mathbf{p}(t)) \tag{102}$$

where we have used the natural pairing $\langle \cdot, \cdot \rangle$ in the place of the dot product used in Definition 1.18.

Corollary 2.2. $\dot{H} = 0$ *along all integral curves of* X_H*.*

Definition 2.8 (Poisson Manifold)**.** *Let* P *be a manifold and let* $\{\cdot, \cdot\}$ *be a Poisson bracket on* P*. The pair* $(P, \{\cdot, \cdot\})$ *is called a* ***Poisson manifold****. Given any smooth* $H : P \to \mathbb{R}$*, the Hamiltonian vector field* X_H *is uniquely determined by the relation*

$$\dot{F} = \mathcal{L}_{X_H} F = \{F, H\} \tag{103}$$

along all integral curves of X_H*.*

Remark that the preceding definition extends Hamiltonian systems to spaces which are not co-tangent bundles.

2.3 Cyclic Coordinates and Symmetries

This section will be somewhat heuristic, but it is important to build intuition before generalizing further. It is sometimes the case when solving problems in Lagrangian mechanics that one notices some coordinates from the domain on which L is defined which are absent from L itself. A good example of this is the Lagrangian for a point particle in the absence of any external potential:

$$L = \frac{1}{2}m||\dot{\mathbf{q}}||^2 \tag{104}$$

In this case, L is entirely independent of $\mathbf{q}$, and we say that the components of $\mathbf{q}$ are cyclic:

Definition 2.9 (Cyclic coordinates)**.** *A coordinate q^i is **cyclic** if L is independent of q^i.*

A consequence of this definition is that, if q^i is a cyclic coordinate of a Lagrangian L, then $\partial L/\partial \dot{q}^i$ is a conserved quantity, i.e. it does not change in time. In the above example we have

$$\frac{d}{dt}\frac{\partial}{\partial \dot{\mathbf{q}}}\left(\frac{1}{2}m||\dot{\mathbf{q}}||^2\right) = \frac{d}{dt}m||\dot{\mathbf{q}}|| = 0 \tag{105}$$

The quantity $m||\dot{\mathbf{q}}||$ corresponds to linear momentum. Note that

$$\frac{\partial L}{\partial \dot{\mathbf{q}}} = m||\dot{\mathbf{q}}|| = \mathbf{p} \tag{106}$$

therefore if a quantity is conserved in the Lagrangian picture of mechanics, it is also conserved in the Hamiltonian picture.

This prompts an interesting exercise: suppose we were to apply a translation $\varphi : \mathbf{q} \mapsto \mathbf{q} + \delta$ before deriving the Euler-Lagrange equations. This transformation is distinct from a change of coordinates because we are still differentiating with respect to $\mathbf{q}$. However, since L is independent of $\mathbf{q}$, we still have

$$\frac{d}{dt}\left(\frac{\partial}{\partial \dot{\mathbf{q}}}(L \circ \varphi)\right) - \frac{\partial}{\partial \mathbf{q}}(L \circ \varphi) = \frac{d}{dt}\left(m||\dot{\mathbf{q}}||\right) = 0 \tag{107}$$

In other words, the Lagrangian does not change under translation, and so the Euler-Lagrange equations one gets from their composition are the same as before the translation was carried out. In light of this, we define a symmetry of a function defined on a manifold M as follows:

Definition 2.10 (Symmetry transformation)**.** *A **symmetry transformation** (or a symmetry) of a function $F : M \to R$ is a differentiable map $\varphi : M \to M$ such that $F \circ \varphi = F$. The function F is **invariant** with respect to the flow Φ if each time-t map Φ_t is a symmetry of F, that is, $F \circ \Phi_t = F$ for all t.*

However, the concept of symmetries obtained via cyclic variables is limited by its dependence on the choice of a particular coordinate system. The rest of this paper is dedicated to developing a coordinate-free concept of an invariance property, and applying it to a particular system.

3 Lie Groups and Symmetry

"Symmetry, as wide or narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection."

- Hermann Weyl, Symmetry (1952)

"Madam, I'm Adam. And Able was I ere I saw Elba."

- James Joyce, Ulysses (1922)

3.1 Lie Groups and Lie Algebras

In the previous section, we encountered symmetries of mechanical systems, defined as transformations that leave a system's behavior unchanged. Mathematically, these symmetries form continuous groups; they can be continuously parameterized, like angles for rotation, or displacements for translation. Lie groups provide the mathematical language for describing these continuous symmetries.

Definition 3.1 (Lie group)**.** *A **Lie group** is a smooth manifold that is also a group, with the property that the binary operation on the group and inversion are smooth.*

Example 3.1. *The group $GL(n)$ of all $n \times n$ invertible matrices is a Lie group. It is an open subset of $\mathcal{M}(n, \mathbb{R})$, and so is an n^2-dimensional submanifold of $\mathcal{M}(n, \mathbb{R})$.*

Example 3.2. *The group $SU(1)$ of all unit-length $z \in \mathbb{C}$ is a Lie group. It is isomorphic to S^1.*

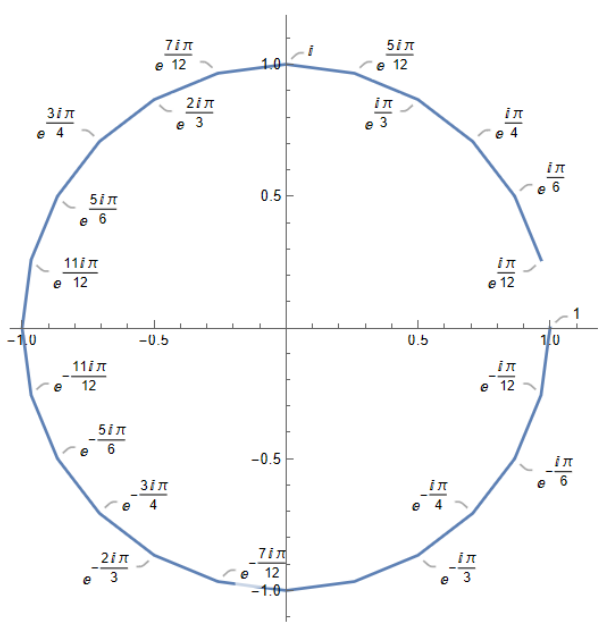


Figure 3: A plot of all complex numbers of the form $z = \exp[i\pi \cdot k/12]$ for $k = 1, \ldots, 24$. Increasing the denominator of the exponential creates a better and better approximation of the unit circle.

Lie groups are fundamental in that, given a local coordinate chart, one may construct an entire set of mutually compatible coordinate charts via binary operations. To illustrate this, take the left translation by $g \in G$ to be the map

$$L_g : G \to G,\ L_g(h) := gh \tag{108}$$

Remark that, by the properties of the Lie group, L_g and L_g^{-1} are smooth. For any chart U which covers the identity, $L_g(U)$ is a chart which covers g. Taking the union over $g \in G$, we find that

$$\bigcup_{g\in G} L_g(U) = G \tag{109}$$

One might instead use the right translation map

$$R_g : G \to G,\ R_g(h) := hg \tag{110}$$

however, we will come to see that left translations provide a natural way to describe Lie group actions. We can also extend left translation to elements of the tangent space. Recall that, for $f : G \to H$, the definition of the tangent map is given by

$$T_p f : T_p G \to T_{f(p)} H \tag{111}$$

Likewise, the left extension of any $\xi \in T_e G$ is the vector field X_ξ^L given by

$$X_\xi^L(g) = T_e L_g(\xi) \tag{112}$$

The reason why we specify $\xi \in T_e G$ will become clear in a moment. For now, we note that there exist some vector fields defined on G which remain unchanged by left multiplication by any group element.

Definition 3.2 (Left invariance)**.** *A vector field $X : G \to TG$, $h \mapsto X(h)$ is called **left invariant** if it is invariant under local left extension, or*

$$T_h L_g(X(h)) = X(L_g(h)) = X(gh) \tag{113}$$

Using a more compact notation, we may also write

$$L_g^*(X) = X \tag{114}$$

Intuitively, we say that the structure of X is preserved over G. A result that is easy to show is that a vector field X defined on G is left invariant if and only if it is the left extension of some $\xi \in T_e G$. If we denote the set of all left invariant vector fields as $\mathfrak{X}_L(G)$, we have

$$\mathfrak{X}_L(G) = \{X_\xi^L : \xi \in T_e G\} \tag{115}$$

Given the existence of left invariant tangent structures on a Lie group G, a natural question to ask is whether G possesses a trivial tangent bundle structure, i.e. that the tangent bundle of the group is isomorphic to the trivial vector bundle $G \times \mathbb{R}^n$. To show that this is indeed the case, we introduce the idea of a Lie algebra, a linear structure that encodes the infinitesimal symmetries of a Lie group. Throughout this paper we will restrict ourselves to Lie algebras over the Reals, although complex Lie algebras exist.

Definition 3.3 (Lie algebra)**.** *A **Lie algebra** is a vector space $\mathfrak{g}$ together with a bilinear operation $(v, w) \in \mathfrak{g} \mapsto [v, w] \in \mathfrak{g}$ called the bracket, such that*

1. $[v, w] = -[w, v]$ *(skew symmetry)*
2. $[[v, w], u] + [[u, v], w] + [[w, u], v] = 0$ *(Jacobi Identity)*

A subspace of $\mathfrak{g}$ which is closed under the bracket is called a Lie subalgebra.

Example 3.3. *The vector space $\mathbb{R}^3$ is a Lie Algebra when equipped with the usual vector cross-product $[\mathbf{x}, \mathbf{y}] = \mathbf{x} \times \mathbf{y}$.*

Example 3.4. *A vector subspace of $\mathcal{M}(n,\mathbb{R})$ with the usual operations of addition and scalar multiplication, equipped with the* ***matrix commutator*** $[A,B]=AB-BA$*, is a Lie algebra.*

Example 3.5. *The vector space $\mathfrak{X}(M)$ of all smooth vector fields on a smooth manifold M is a Lie algebra. The bracket is the* ***Jacobi-Lie bracket****,*

$$[X,Y]=(DY)\cdot X-(DX)\cdot Y \tag{116}$$

Given the sheer number of Lie algebra structures from which to choose, we look for a way to assign to a Lie group G with a canonical basis a *unique* Lie algebra $\mathfrak{g}$. This Lie algebra should emphasize the trivial structure of the vector bundle on G. To this end, we prove the following two Lemmas:

Lemma 3.1. *Let $\mathfrak{X}_L(G)$ be the set of all left invariant vector fields on a Lie group G, and let $\mathfrak{X}(G)$ be the set of all smooth vector fields on G. Then*

1. *$\mathfrak{X}_L(G)$ is a vector subspace of $\mathfrak{X}(G)$.*
2. *$\mathfrak{X}_L(G)$ is a Lie subalgebra of $\mathfrak{X}(G)$.*

Proof. The first condition is trivial. For the second, We need to show that $\mathfrak{X}_L(G)$ is closed under the bracket of $\mathfrak{X}(G)$. We compute, for any $\xi,\eta\in T_eG$,

$$[X_\xi^L,X_\eta^L]=[L_g^*X_\xi^L,L_g^*X_\eta^L]=L_g^*[X_\xi^L,X_\eta^L] \tag{117}$$

i.e. $[X_\xi^L,X_\eta^L]$ is left-invariant. □

Lemma 3.2. *The map λ, defined as*

$$\lambda:T_eG\to\mathfrak{X}_L(G),\ \xi\mapsto X_\xi^L \tag{118}$$

is a vector space isomorphism.

Proof. We check, for any $\xi,\ \eta\in T_eG$,

$$\lambda(a\xi+b\eta)=X_{a\xi+b\eta}^L=T_eL_g(a\xi+b\eta)=aT_eL_g(\xi)+bT_eL_g(\eta)=a\lambda(\xi)+b\lambda(\eta) \tag{119}$$

where in the second-to-last step we have used the linearity of $T_eL_g(\xi)$. As for the existence of λ^{-1}, the only issue we may run into is that λ does not have a trivial kernel. Assume this is the case, i.e. that there exist ξ_1 and ξ_2 in T_eG such that $\xi_1\neq\xi_2$, but $\lambda(\xi_1)=\lambda(\xi_1)=X_0^L$. Then by the linearity of λ we have

$$\lambda(\xi_1)-\lambda(\xi_2)=\lambda(\xi_1-\xi_2)=0 \tag{120}$$

which implies

$$X_{\xi_1-\xi_2}^L=T_eL_g(\xi_1-\xi_2)=0 \tag{121}$$

for all $g\in G$. Since the tangent map of L_g is linear and invertible, it must be that $\xi_1=\xi_2$. □

These two results justify the following definition:

Definition 3.4 (Lie algebra of a Lie group G)**.** *Let G be a Lie group. The tangent space at the identity T_eG of G, together with the Lie bracket defined by the isomorphism λ,*

$$[\xi,\eta]:=[X_\xi^L,X_\eta^L](e) \tag{122}$$

is called the Lie algebra of G.

3.1.1 Vector Representations of Lie Algebras

Similar to how a group isomorphism preserves the binary operation, a Lie algebra isomorphism preserves the bracket structure. These kinds of mappings are useful when constructing vector representations of elements of a Lie algebra, which in turn are useful when trying to perform computations.

Definition 3.5 (Lie algebra isomorphism)**.** *Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras. A linear map $\rho : \mathfrak{g} \to \mathfrak{h}$ is called a **Lie algebra isomorphism** if*

1. *$\rho([\xi, \eta]) = [\rho(\xi), \rho(\eta)]$, for all $\xi, \eta \in \mathfrak{g}$.*
2. *ρ is bijective.*

Example 3.6 (Vector representation of $\mathfrak{so}(2)$)**.** *The Lie algebra $\mathfrak{so}(2)$ is isomorphic to the real line, with*

$$\rho : \mathbb{R} \to \mathfrak{so}(2), \quad \xi \mapsto \begin{bmatrix} 0 & -\xi \\ \xi & 0 \end{bmatrix} \tag{123}$$

The commutator of any two elements of $\mathfrak{so}(2)$ is always zero, so if we define the bracket on $\mathbb{R}$ to be zero, then this map is both linear and bracket-preserving.

Example 3.7 (Vector representation of $\mathfrak{so}(3)$)**.** *The Lie algebra $\mathfrak{so}(3)$ can be identified with the vector space $\mathbb{R}^3$ via the hat map, defined by*

$$\widehat{(\dots)} : \mathbb{R}^3 \to \mathfrak{so}(3), \quad \mathbf{x} = (x^1, x^2, x^3) \mapsto \hat{\mathbf{x}} = \begin{bmatrix} 0 & -x^3 & x^2 \\ x^3 & 0 & -x^1 \\ -x^2 & x^1 & 0 \end{bmatrix} \tag{124}$$

It is easy to show that, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, we have

$$\hat{\mathbf{x}}\mathbf{y} = \mathbf{x} \times \mathbf{y} \tag{125}$$

3.2 The Exponential Map

We have shown that, given a Lie group G, one can easily construct a corresponding Lie algebra $\mathfrak{g}$. Now we examine the inverse problem: given elements from $\mathfrak{g}$, we want to recapture the local group structure of G. To accomplish this, we turn to the exponential map.

To begin, let $\xi \in \mathfrak{g}$ and consider $X_\xi^L = T_e L_g(\xi)$. The **one-parameter subgroup** corresponding to ξ, denoted $\gamma_\xi : \mathbb{R} \to G$, is the unique integral curve which solves the initial value problem

$$\frac{dg}{dt} = X_\xi^L(g) \tag{126}$$

$$g(0) = e \tag{127}$$

One can show that the image of this curve is a subgroup of G. Given this definition of γ_ξ, one can in turn define the **exponential map** on G to be the unique function satisfying

$$\exp : \mathfrak{g} \to G, \ \xi \mapsto \exp(\xi) := \gamma_\xi(1) \tag{128}$$

Example 3.8 (Matrix Exponential)**.** *Let $A \in \mathfrak{gl}(n) = \mathcal{M}(n, \mathbb{R})$ and consider the usual exponential of matrices*

$$e^{tA} = \sum_{n=0}^{\infty} \frac{1}{n!} t^n A^n \tag{129}$$

If we define $\gamma_A(t) = e^{tA}$, it is easy to show that $\gamma_A(0) = \mathbb{I}_n$. It is also easy to show that $d/dt(e^{tA}) = e^{tA}A$, and using the definition of the left extension, we arrive at

$$\frac{d}{dt} e^{tA} = X_A(e^{tA}) \tag{130}$$

Therefore we conclude that $\gamma_A(t)$ is the one parameter subgroup generated by A, with

$$exp(A) = \gamma_A(1) = e^A \tag{131}$$

Proposition 3.1 (Properties of the exponential map)**.** *The following statements hold:*

1. $exp(t\xi) = \gamma_\xi(t) \quad \forall t \in \mathbb{R}$
2. $\gamma_\xi(s+t) = \gamma_\xi(s)\gamma_\xi(t) \quad \forall s, t \in \mathbb{R}$
3. *All smooth one-parameter subgroups of G are of the form $\{exp(t\xi) : t \in \mathbb{R}\}$ for some $\xi \in \mathfrak{g}$*
4. *The exponential map is a local C^∞ diffeomorphism from a neighbourhood of $0 \in \mathfrak{g}$ onto a neighbourhood of $e \in G$*

Corollary 3.1. *The exponential map induces a coordinate chart in a neighborhood of e. The coordinates associated to this chart are called **canonical coordinates** of the Lie group G.*

3.3 Lie Group Actions

Lie group actions generalize flows. An action of a group on a set is a map that associates to each element of the group an invertible transformation, usually a diffeomorphism, of the given set, in such a way that the group operation corresponds to composition of transformations. Thus, the group may be thought of as a group of transformations. If the diffeomorphisms corresponding to the group elements all leave a certain function invariant, then the group (with the specified action) is a symmetry group of that function.

Definition 3.6 (Left action)**.** *A **left action** of a Lie group G on a manifold M is a smooth map $\Phi : G \times M \to M$ such that*

1. *$\Phi(e, x) = x$ for all $x \in M$*
2. *$\Phi(g, \Phi(h, x)) = \Phi(gh, x)$ for all $g, h \in G$, $x \in M$*
3. *For every $g \in G$, the map $\Phi_g : M \to M$ defined as*

$$\Phi_g(x) := \Phi(g, x) \tag{132}$$

is a diffeomorphism.

Example 3.9. *The standard action of a matrix Lie group $G \subset GL(n, \mathbb{R})$ on $\mathbb{R}^n$ is given by*

$$\Phi(A, \mathbf{v}) = A\mathbf{v} \tag{133}$$

Of course, the above example uses a group action which is generated explicitly by an element g in G. This is not always necessary: using the exponential map, we can reformulate group actions on a set in the following way:

Definition 3.7 (Infinitesimal generator)**.** *Consider the left action Φ of a Lie group G on the manifold M such that $\mathbf{x} \mapsto g\mathbf{x}$. Let $\xi \in \mathfrak{g}$ be a vector in the Lie algebra of G and consider the one-parameter sub-group $\{\exp(t\xi) : t \in \mathbb{R}\} \subset G$. The orbit of $\mathbf{x} \in M$ with respect to this subgroup is a smooth path in M. The **infinitesimal generator** associated to ξ at $\mathbf{x} \in M$, denoted $\xi_M(\mathbf{x})$, is the tangent vector to this path at $\mathbf{x}$, i.e.*

$$\xi_M(\mathbf{x}) := \left.\frac{d}{dt}\right|_{t=0} (\exp(t\xi)\mathbf{x}) \in T_{\mathbf{x}}M \tag{134}$$

The smooth vector field assigning to each $\mathbf{x} \in M$ its corresponding infinitesimal generator associated to ξ is called the infinitesimal generator vector field associated to ξ.

Example 3.10 (The infinitesimal generator for the $SO(2)$ action on $\mathbb{R}^2$)**.** *Let*

$$\xi = \begin{bmatrix} 0 & -\xi \\ \xi & 0 \end{bmatrix} \in \mathfrak{so}(2), \quad \mathbf{x} = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \in \mathbb{R}^2 \tag{135}$$

Then

$$\begin{aligned} \xi_{\mathbb{R}^2}(\mathbf{x}) &= \left.\frac{d}{dt}\right|_{t=0} (\exp(t\xi)\mathbf{x}) && (136) \\ &= \left.\frac{d}{dt}\right|_{t=0} \left(\sum_n \frac{1}{n!} t^n \xi^n \right) \mathbf{x} && (137) \\ &= (\xi + \mathcal{O}(t))\,\mathbf{x} && (138) \\ &= \xi \mathbf{x} && (139) \\ &= \begin{bmatrix} -\xi x^2 \\ \xi x^1 \end{bmatrix} && (140) \end{aligned}$$

The intuition for the definition of an infinitesimal generator is that the exponential map "projects down" from the Lie algebra onto its corresponding group element, which then acts on $\mathbf{x}$. By taking the time derivative of the map, we obtain a tangent vector which encodes the change in $\mathbf{x}$ brought on by g. In this way, one can describe a group action on a manifold by expressing the magnitude and direction of the changes brought on by the action.

Finally, we introduce a definition which generalizes Definition 2.10:

Definition 3.8 (Invariance)**.** *A function F is* ***invariant*** *with respect to an action Φ of a Lie group G if, for every $g \in G$, the map Φ_g is a symmetry of F , that is, $F \circ \Phi_g = F$. The group G is called a Lie group symmetry or* ***symmetry group*** *of F.*

This relates to the following result which is useful for our purposes:

Proposition 3.2. *Let $\Phi : G \times Q \to Q$ be a flow. Let $L : TQ \to \mathbb{R}$ be a hyperregular Lagrangian, with corresponding Hamiltonian $H : T^*Q \to \mathbb{R}$. If L is invariant with respect to the tangent lift of Φ then H is invariant with respect to the cotangent lift of Φ.*

3.3.1 The Adjoint and Co-adjoint Actions

A Lie group's action extends not only to generic manifolds, but to members of its Lie algebra. Since these are vectors, we are motivated to come up with linear transformations based on group actions. The final definition of the previous section is informative here: it will be important to understand the action of a group G on itself when we begin to define functions on G or $\mathfrak{g}$ which are invariant with respect to an action.

Lie group actions can be lifted to actions on the tangent bundle TG and the cotangent bundle T^*G. For actions on the tangent bundle, we have

$$\begin{aligned} G \times TG &\to TG && (141) \\ (g, (h, v)) \mapsto (gh, gv) := (gh, T_h L_g(v)) &= \left(gh, \left.\frac{d}{dt}\right|_{t=0} (gc(t)) \right) && (142) \end{aligned}$$

where $c(t)$ is any path in G with $c(0) = h$ and $c'(0) = v$, and

$$\begin{aligned} G \times T^*G &\to T^*G && (143) \\ (g, (h, \alpha)) \mapsto (gh, g\alpha) &:= \left(gh, T^*_{gh} L_{g^{-1}}(\alpha)\right) && (144) \end{aligned}$$

where

$$\langle T^*_{gh} L_{g^{-1}}(\alpha), w\rangle = \langle \alpha, T_{gh} L_{g^{-1}}(w)\rangle \quad \forall w \in T_{gh}G \tag{145}$$

The above notation looks somewhat formidable, so we introduce some simplifying notation:

Definition 3.9 (Conjugation)**.** *The **conjugation** of an element $h \in G$ by $g \in G$ is defined as*

$$I_g : G \times G \to G,\ I_g(h) := ghg^{-1} \tag{146}$$

We say that two elements $a, b \in G$ are conjugate if there exists $g \in G$ such that $b = gag^{-1}$. Orbits of this action are called conjugacy classes.

Note that ghg^{-1} can also be written as $(L_g \circ R_{g^{-1}})(h)$.

Definition 3.10 (Adjoint action, co-adjoint action)**.** *The **adjoint action** of G on $\mathfrak{g}$ is defined by*

$$Ad_g : G \times \mathfrak{g} \to \mathfrak{g},\ Ad_g(\xi) := T_e I_g(\xi) = T_e(L_g \circ R_{g^{-1}})(\xi) \tag{147}$$

*The **co-adjoint action** of G on $\mathfrak{g}^*$ is defined as the inverse dual to the Adjoint action, i.e.*

$$Ad^*_{g^{-1}} : G \times \mathfrak{g}^* \to \mathfrak{g}^* \tag{148}$$

where

$$\langle Ad^*_{g^{-1}} \mu, \xi\rangle = \langle \mu, Ad_{g^{-1}}\xi\rangle \tag{149}$$

Example 3.11 (The adjoint action for matrix Lie groups)**.** *Let G be any matrix Lie group, i.e. $G \subseteq \mathcal{M}(n, \mathbb{R})$. Let $B : I \subset \mathbb{R} \to G$ be a path in G such that $B(0) = \mathbb{I}_n$ and $B'(0) = \xi \in \mathfrak{g}$. Then, given $g = \mathbf{M} \in G$, we have*

$$\begin{aligned} Ad_g(\xi) &= Ad_{\mathbf{M}}(\xi) && (150)\\ &= T_{\mathbb{I}_n}(L_{\mathbf{M}} \circ R_{\mathbf{M}^{-1}})(\xi) && (151)\\ &= \left.\frac{d}{dt}\right|_{t=0} \mathbf{M}B(t)\mathbf{M}^{-1} && (152)\\ &= \mathbf{M}\xi\mathbf{M}^{-1} && (153) \end{aligned}$$

We can use this formulation of Ad_g to identify the co-adjoint action: let $\mu \in \mathfrak{g}^*$ and take

$$\langle Ad^*_{\mathbf{M}^{-1}} \mu, \xi\rangle = \langle \mu, Ad_{\mathbf{M}^{-1}}\xi\rangle \tag{154}$$

Using our expression for the adjoint action, and recalling that the trace pairing $\langle \cdot, \cdot\rangle : \mathcal{M}(n, \mathbb{R}) \times \mathcal{M}(n, \mathbb{R}) \to \mathbb{R}$, $\langle \mathbf{C}, \mathbf{D}\rangle = \sum_{i,j} C_{ij} D_{ij} = tr(\mathbf{C}^T\mathbf{D})$, we have

$$\begin{aligned} \langle \mu, \mathbf{M}^{-1}\xi\mathbf{M}\rangle &= tr(\mu\mathbf{M}^T\xi^T\mathbf{M}^{-T}) && (155)\\ &= tr(\mathbf{M}^{-T}\mu\mathbf{M}^T\xi^T) && (156)\\ &= \langle \mathbf{M}^{-T}\mu\mathbf{M}^T, \xi\rangle && (157) \end{aligned}$$

Therefore,

$$Ad^*_{\mathbf{M}^{-1}}\mu = \mathbf{M}^{-T}\mu\mathbf{M}^T \tag{158}$$

One can also define infinitesimal generators at $g \in G$. The infinitesimal generator of the adjoint action on G is the vector field ξ_g defined by

$$\xi_g(\eta) = \left.\frac{d}{dt}\right|_{t=0} Ad_{\exp(t\xi)}\eta = T_e(Ad_g\eta)\xi \tag{159}$$

Since $\mathfrak{g}$ is a vector space, $T\mathfrak{g} = \mathfrak{g} \times \mathfrak{g}$, so infinitesimal generators can be identified with maps from $\mathfrak{g}$ onto itself.

Definition 3.11 (Adjoint action of $\mathfrak{g}$, adjoint operator)**.** *The infinitesimal generator map*

$$ad_\xi(\eta) := \xi_g(\eta) = \frac{d}{dt}\bigg|_{t=0} Ad_{\exp(t\xi)}\eta \tag{160}$$

*is called the **adjoint action of** $\mathfrak{g}$ on itself. The **adjoint operator** on $\mathfrak{g}$ is denoted by ad.*

Example 3.12 (The adjoint action for matrix Lie algebras)**.** *Recall from Example 3.11 that, for any Lie group $G \subset \mathcal{M}(n, \mathbb{R})$,*

$$Ad_{\mathbf{M}}\xi = \mathbf{M}\xi\mathbf{M}^{-1} \tag{161}$$

It follows that, for any $\eta \in \mathfrak{g}$, we have

$$\begin{aligned} ad_\xi\eta &= \frac{d}{dt}\bigg|_{t=0} Ad_{\exp(t\xi)}\eta && (162)\\ &= \frac{d}{dt}\bigg|_{t=0} \exp(t\xi)\eta\exp(-t\xi) && (163)\\ &= \xi\eta - \eta\xi = [\xi, \eta] && (164)\end{aligned}$$

This result does not only hold for matrix Lie algebras:

Proposition 3.3. *For any Lie algebra $\mathfrak{g}$ and for any $\xi, \eta \in \mathfrak{g}$, we have*

$$ad_\xi\eta = [\xi, \eta] \tag{165}$$

Proof.

$$\begin{aligned} \mathrm{ad}_\xi\eta &= \frac{d}{dt}\bigg|_{t=0} T_e(L_{\exp(t\xi)} \circ R_{-\exp(t\xi)})\eta && (166)\\ &= \frac{d}{dt}\bigg|_{t=0} T_{\exp(t\xi)}R_{-\exp(t\xi)}(T_eL_{\exp(t\xi)}(\eta)) && (167)\\ &= \frac{d}{dt}\bigg|_{t=0} T_{\exp(t\xi)}R_{-\exp(t\xi)}(X^L_\eta(\exp t\xi)) && (168)\\ & && (169)\end{aligned}$$

Using $\exp t\xi = e \cdot \exp t\xi = R_{\exp t\xi} \cdot e = \Phi_t(e)$, we have

$$\begin{aligned} \mathrm{ad}_\xi\eta &= \frac{d}{dt}\bigg|_{t=0} T_{\Phi_t(e)}\Phi_t^{-1}(X^L_\eta(\Phi_t(e))) && (170)\\ &= \frac{d}{dt}\bigg|_{t=0} (\Phi_t^* X^L_\eta)(e) && (171)\\ &= \mathcal{L}_{X^L_\xi} X^L_\eta(e) && (172)\\ &= [X^L_\xi, X^L_\eta](e) = [\xi, \eta] && (173)\end{aligned}$$

□

Definition 3.12 (Co-adjoint operator)**.** *The **co-adjoint operator** is the map*

$$\begin{aligned} ad^* : \mathfrak{g} \times \mathfrak{g}^* &\to \mathfrak{g}^* && (174)\\ (\xi, \mu) &\mapsto ad^*_\xi\mu && (175)\end{aligned}$$

such that

$$\langle ad^*_\xi\mu, \eta\rangle = \langle\mu, ad_\xi\eta\rangle \tag{176}$$

Remark*: the infinitesimal generator of the co-adjoint action is*

$$\xi_{\mathfrak{g}^*}(\mu) = -ad^*_\xi(\mu) \tag{177}$$

Example 3.13 (The co-adjoint action for matrix Lie algebras)**.** *From the above definition we have*

$$\langle ad^*_\xi \mu, \eta\rangle = \langle \mu, ad_\xi \eta\rangle = \langle \mu, \xi\eta - \eta\xi\rangle = \langle \mu, \xi\eta\rangle - \langle \mu, \eta\xi\rangle \tag{178}$$

Using the trace pairing, we see that

$$\begin{aligned}
\langle \mu, \xi\eta\rangle - \langle \mu, \eta\xi\rangle &= tr(\mu(\xi\eta)^T) - tr(\mu(\eta\xi)^T) && (179)\\
&= tr(\mu\eta^T\xi^T) - tr(\mu\xi^T\eta^T) && (180)\\
&= tr(\xi^T\mu\eta^T) - tr(\mu\xi^T\eta^T) && (181)\\
&= tr((\xi^T\mu - \mu\xi^T)\eta^T) && (182)\\
&= \langle -[\mu, \xi^T], \eta\rangle && (183)
\end{aligned}$$

from which we identify

$$ad^*_\xi \mu = -[\mu, \xi^T] \tag{184}$$

If $G = SO(n)$, we have that $\xi = -\xi^T$ for all $\xi \in \mathfrak{so}(n)$, and so

$$ad^*_\xi \mu = [\mu, \xi] \qquad in\, \mathfrak{so}(n) \tag{185}$$

The adjoint and co-adjoint actions on $\mathfrak{g}$ and $\mathfrak{g}^*$, respectively, are the linear transformations mentioned at the beginning of this section. We now have all the tools we need to derive and begin studying reductions.

4 Reductions

"From a drop of water, a logician could infer the possibility of an Atlantic or a Niagara without having seen or heard of one or the other."

- Arthur Conan Doyle, A Study in Scarlet (1887)

We have seen how the symmetries of a Lagrangian or Hamiltonian system defined on a Lie group G can help us find quantities inherent to the system which remain constant over time. Now, we study applications of those symmetries, exploiting them to reduce the computational complexity of finding equations of motion. This process of shaving off redundant quantities from a mechanical system is known as a reduction, and we will carry it out in two examples: one using the Lagrangian formalism, and one using the Hamiltonian formalism.

4.1 Rigid Body Dynamics

Consider a rigid body with a fixed point. It is usually assumed that this fixed point is the centre of mass of the body, but this is not necessary, and will not be true when we study the heavy symmetric top. Given a reference configuration of the body, two systems of coordinates are introduced: a fixed inertial **spatial coordinate system**, and a moving **body coordinate system**, both with origin at the fixed point of the body.

The position of a particle in body coordinates is called the particle's **label**. The configuration of the body at time t is determined by a rotation matrix $\mathbf{R}(t)$ that takes the label $\mathbf{X}$ of any particle in the body to its

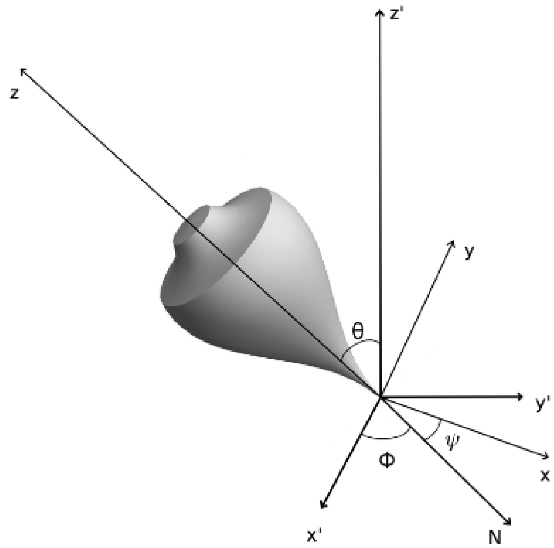


Figure 4: Symmetric top with fixed point [4]

spatial position $\mathbf{x}(t)$. From this, we determine that the configuration space of the rigid body is $SO(3)$, and $\mathbf{R}(t)$ is a path in $SO(3)$. The position and velocity at time t of the particle with label $\mathbf{X}$ are given as

$$\mathbf{x}(t) = \mathbf{R}(t)\mathbf{X},\ \dot{\mathbf{x}}(t) = \dot{\mathbf{R}}(t)\mathbf{X} = \dot{\mathbf{R}}(t)\mathbf{R}^{-1}(t)\mathbf{x}(t) \tag{186}$$

Proposition 4.1. *For any* $\mathbf{R} \in SO(3)$, $\dot{\mathbf{R}}\mathbf{R}^{-1}$ *and* $\mathbf{R}^{-1}\dot{\mathbf{R}}$ *are elements of* $\mathfrak{so}(3)$.

Proof. Since $\mathfrak{so}(3) = T_{\mathbf{I}}SO(3)$, any tangent vector $(\mathbf{R}, \dot{\mathbf{R}}) \in TSO(3)$ can be translated to $\mathfrak{so}(3)$ by either left or right multiplication by $\mathbf{R}^{-1}$:

$$TL_{\mathbf{R}^{-1}}(\mathbf{R}, \dot{\mathbf{R}}) = (\mathbf{I}, \mathbf{R}^{-1}\dot{\mathbf{R}}),\ TR_{\mathbf{R}^{-1}}(\mathbf{R}, \dot{\mathbf{R}}) = (\mathbf{I}, \dot{\mathbf{R}}\mathbf{R}^{-1}) \tag{187}$$

□

Note that the above result reveals that $\dot{\mathbf{R}}\mathbf{R}^{-1}$ and $\mathbf{R}^{-1}\dot{\mathbf{R}}$ are skew-symmetric.

We define the **spatial angular velocity** vector ω as

$$\widehat{\omega} = \dot{\mathbf{R}}\mathbf{R}^{-1} \tag{188}$$

where

$$\dot{\mathbf{x}} = \dot{\mathbf{R}}\mathbf{R}^{-1}\mathbf{x} = \widehat{\omega}\mathbf{x} = \omega \times \mathbf{x} \tag{189}$$

We also define the **body angular velocity** vector $\mathbf{\Omega}$ as

$$\mathbf{\Omega} = \mathbf{R}^{-1}\omega \tag{190}$$

A straightforward calculation shows us that

$$\mathbf{\Omega} \times \mathbf{X} = \mathbf{R}^{-1}\omega \times \mathbf{R}^{-1}\mathbf{x} = \mathbf{R}^{-1}(\omega \times \mathbf{x}) = \mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{R}^{-1}\mathbf{x} = \mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{X} \tag{191}$$

which gives

$$\widehat{\mathbf{\Omega}} = \mathbf{R}^{-1}\dot{\mathbf{R}} \tag{192}$$

Let $\rho(\mathbf{X})$ be the density of the body at position $\mathbf{X}$ in body coordinates, which we assume to be constant. Let $\mathcal{B}$ be the region occupied by the body at time $t = 0$. Then the mass of the body is given by

$$m = \int_{\mathcal{B}} \rho(\mathbf{X}) d^3\mathbf{X} \tag{193}$$

We define the **kinetic energy** of the rigid body as

$$K = \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\dot{\mathbf{x}}||^2 d^3\mathbf{X} \tag{194}$$

We can massage this equation to obtain an expression fully in terms of the label $\mathbf{X}$:

$$\begin{align}
\frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\dot{\mathbf{x}}||^2 d^3\mathbf{X} &= \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\dot{\mathbf{R}}\mathbf{X}||^2 d^3\mathbf{X} \tag{195}\\
&= \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{X}||^2 d^3\mathbf{X} \tag{196}\\
&= \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\widehat{\mathbf{\Omega}}\mathbf{X}||^2 d^3\mathbf{X} \tag{197}\\
& \tag{198}
\end{align}$$

If we identify $\widehat{\mathbf{\Omega}}\mathbf{X}$ with $-\widehat{\mathbf{X}}\mathbf{\Omega}$, we obtain

$$K = \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\widehat{\mathbf{X}}\mathbf{\Omega}||^2 d^3\mathbf{X} = \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})\mathbf{\Omega}^T\widehat{\mathbf{X}}^T\widehat{\mathbf{X}}\mathbf{\Omega} d^3\mathbf{X} = \frac{1}{2}\mathbf{\Omega}^T\left(\int_{\mathcal{B}} \rho(\mathbf{X})\widehat{\mathbf{X}}^T\widehat{\mathbf{X}} d^3\mathbf{X}\right)\mathbf{\Omega} \tag{199}$$

Identifying $\widehat{\mathbf{X}}^T$ with $-\widehat{\mathbf{X}}$ and expanding the integrand gives

$$\widehat{\mathbf{X}}^T\widehat{\mathbf{X}} = -\widehat{\mathbf{X}}^2 = \begin{bmatrix} (X^2)^2+(X^3)^2 & -(X^1)(X^2) & -(X^1)(X^3) \\ -(X^1)(X^2) & (X^1)^2+(X^3)^2 & -(X^2)(X^3) \\ -(X^1)(X^3) & -(X^2)(X^3) & (X^1)^2+(X^2)^2 \end{bmatrix} \tag{200}$$

Which is equivalent to $||\mathbf{X}||^2\mathbb{I}_3 - \mathbf{X}\mathbf{X}^T$. We denote the integral in Equation 199 the **moment of inertia tensor**,

$$\mathbb{I} := \int_{\mathcal{B}} \rho(\mathbf{X})\left(||\mathbf{X}||^2\mathbb{I}_3 - \mathbf{X}\mathbf{X}^T\right) d^3\mathbf{X} \tag{201}$$

Note the use of subscripts to differentiate the 3×3 identity matrix. We can now express the kinetic energy of the rotating body as

$$K = \frac{1}{2}\mathbf{\Omega}^T\mathbb{I}\mathbf{\Omega} \tag{202}$$

4.2 Euler-Poincare Reduction

One objection to the way we have written the kinetic energy term above is that it no longer contains explicit mention of any element of $SO(3)$. In fact, we could have defined the kinetic energy of the system another way:

$$\begin{align}
K &= \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})||\widehat{\mathbf{\Omega}}\mathbf{X}||^2 d^3\mathbf{X} \tag{203}\\
&= \frac{1}{2}\int_{\mathcal{B}} \rho(\mathbf{X})\mathrm{tr}\left((\widehat{\mathbf{\Omega}}\mathbf{X})(\widehat{\mathbf{\Omega}}\mathbf{X})^T\right) d^3\mathbf{X} \tag{204}\\
&= \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{\Omega}}\left(\int_{\mathcal{B}} \rho(\mathbf{X})\mathbf{X}\mathbf{X}^T d^3\mathbf{X}\right)\widehat{\mathbf{\Omega}}^T\right) \tag{205}\\
&= \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{\Omega}}\mathbb{J}\widehat{\mathbf{\Omega}}^T\right) \tag{206}\\
&= \frac{1}{2}\mathrm{tr}\left((\mathbf{R}^{-1}\dot{\mathbf{R}})\mathbb{J}(\mathbf{R}^{-1}\dot{\mathbf{R}})^T\right) \tag{207}\\
&= \frac{1}{2}\mathrm{tr}\left(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T\right) \tag{208}
\end{align}$$

where we define $\mathbb{J} := \int_{\mathcal{B}} \rho(\mathbf{X})\mathbf{X}\mathbf{X}^T d^3\mathbf{X}$ to be the **coefficient of inertia tensor** of the body with respect to the origin. The Lagrangian of the system with respect to $SO(3)$, as well as the corresponding Euler-Lagrange equations, would be given by

$$L(\mathbf{R}, \dot{\mathbf{R}}) = \frac{1}{2}\mathrm{tr}\left(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T\right) - V(\mathbf{R}), \tag{209}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{R}}}\right) - \frac{\partial L}{\partial \mathbf{R}} = 0 \tag{210}$$

While technically correct, these equations are tricky to work with: we would have to consider $SO(3)$ as a submanifold of $\mathcal{M}(3\mathbb{R}) \cong \mathbb{R}^9$ and model the the evolution of the body using a system of constraint equations. Depending on the form of the potential energy $V(\mathbf{R})$, however, it is sometimes possible to make use of the system's inherent symmetry to "reduce" the dynamics of the system. The Lagrangian of the reduced system will no longer act on TG, but on $\mathfrak{g}$, which is based solely at the identity.

To expand upon this, we prove the follow Lemma.

Lemma 4.1. *Let K be the kinetic energy of a rigid body system as defined in 208. Then K is left invariant with respect to tangent-lifted left translation.*

Proof. Let $\mathbf{Q} \in SO(3)$. We compute

$$\frac{1}{2}\mathrm{tr}\left((\mathbf{Q}\dot{\mathbf{R}})\mathbb{J}(\mathbf{Q}\dot{\mathbf{R}})^T\right) = \frac{1}{2}\mathrm{tr}\left(\mathbf{Q}\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T\mathbf{Q}^T\right) = \frac{1}{2}\mathrm{tr}\left(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T\right) = K \tag{211}$$

□

In the proceeding calculations we will assume left-independence of all our Lagrangians. For this purpose we define the **left-trivialized coordinates**

$$(g, \dot{g}) \mapsto (g, g^{-1}\dot{g}) = (g, T_g L_{g^{-1}}\dot{g}) \tag{212}$$

Left-trivialized coordinates represent a tangent vector at a point on a Lie group in terms of its coordinates in the tangent space at the identity, effectively "flattening" the tangent space at that point into the tangent space at the identity.

Proposition 4.2. *Let G be a Lie group together with a Lagrangian $L : TG \to \mathbb{R}$. In left-trivialized coordinates, the equations of motion are given by*

$$\frac{d}{dt}\left(\frac{\delta L}{\delta \xi}\right) = ad^*_\xi \frac{\delta L}{\delta \xi} + T_e^* L_g \left(\frac{\delta L}{\delta g}\right) \tag{213}$$

Proof. Define the Lagrangian in left-trivialized coordinates; that is,

$$\tilde{L}(g, \xi) = L(g, g\xi) \tag{214}$$

Along these paths, Hamilton's Principle states

$$\delta \int_a^b \tilde{L}(g(t), \xi(t))dt = 0 \tag{215}$$

Using variational derivative notation, one can write

$$\int_a^b \langle \frac{\delta \tilde{L}}{\delta g}, \delta g\rangle + \langle \frac{\delta \tilde{L}}{\delta \xi}, \delta \xi\rangle dt = 0 \tag{216}$$

We will assume that G is a matrix Lie group for the purposes of this proof. Define $g_\varepsilon(t)$ to be a family of curves in G such that $g_0(t) = g(t)$ and let

$$\delta g = \left.\frac{dg_\varepsilon(t)}{d\varepsilon}\right|_{\varepsilon=0} \tag{217}$$

For $\xi = g^{-1}\dot{g}$, the variation is computed as

$$\delta\xi = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}(g_\varepsilon^{-1}\dot{g}_\varepsilon) = -g^{-1}(\delta g)g^{-1}\dot{g} + g^{-1}\frac{d^2g}{dtd\varepsilon}\bigg|_{\varepsilon=0} \tag{218}$$

Set $\eta := g^{-1}\delta g$, i.e. $\eta(t)$ is an arbitrary path in $\mathfrak{g}$ that vanishes at the endpoints. We compute

$$\frac{d\eta}{dt} = -g^{-1}\dot{g}g^{-1}(\delta g) + g^{-1}\frac{d^2g}{dtd\varepsilon}\bigg|_{\varepsilon=0} \tag{219}$$

Taking the difference of the two terms, we find

$$\begin{aligned} \delta\xi - \frac{d\eta}{dt} &= -g^{-1}(\delta g)g^{-1}\dot{g} + g^{-1}\dot{g}g^{-1}(\delta g) && (220)\\ &= \xi\eta - \eta\xi && (221)\\ &= [\xi, \eta] && (222) \end{aligned}$$

Or,

$$\delta\xi = \dot{\eta} + [\xi, \eta] = \dot{\eta} + \mathrm{ad}_\xi \eta \tag{223}$$

Substituting this into the above and noting that $\delta g = g\eta$, we have

$$\begin{aligned} 0 &= \int_a^b \langle \frac{\delta \tilde{L}}{\delta g}, \delta g\rangle + \langle \frac{\delta \tilde{L}}{\delta \xi}, \delta\xi\rangle dt && (224)\\ &= \int_a^b \langle \frac{\delta \tilde{L}}{\delta g}, g\eta\rangle + \langle \frac{\delta \tilde{L}}{\delta \xi}, \dot{\eta} + \mathrm{ad}_\xi\eta\rangle dt && (225)\\ &= \int_a^b \langle \frac{\delta \tilde{L}}{\delta g}, T_eL_g(\eta)\rangle + \langle \frac{\delta \tilde{L}}{\delta \xi}, \frac{d\eta}{dt}\rangle + \langle \frac{\delta \tilde{L}}{\delta \xi}, \mathrm{ad}_\xi\eta\rangle dt && (226)\\ &= \int_a^b \langle T_e^*L_g\left(\frac{\delta \tilde{L}}{\delta g}\right), \eta\rangle + \langle -\frac{d}{dt}\left(\frac{\delta \tilde{L}}{\delta \xi}\right) + \mathrm{ad}_\xi^*\left(\frac{\delta \tilde{L}}{\delta \xi}\right), \eta\rangle dt && (227)\\ &= \int_a^b \langle T_e^*L_g\left(\frac{\delta \tilde{L}}{\delta g}\right) - \frac{d}{dt}\left(\frac{\delta \tilde{L}}{\delta \xi}\right) + \mathrm{ad}_\xi^*\left(\frac{\delta \tilde{L}}{\delta \xi}\right), \eta\rangle dt && (228) \end{aligned}$$

Since the integrand must be zero for any path $\eta(t)$, the result follows. $\square$

Note that, if the original Lagrangian L is left invariant, then

$$\tilde{L}(g, \xi) = L(e, \xi) \tag{229}$$

i.e. L is independent of g, and so the variational derivative $\delta L/\delta g$ vanishes. In this case, we are left with

$$\frac{d}{dt}\left(\frac{\delta l}{\delta \xi}\right) = \mathrm{ad}_\xi^*\frac{\delta l}{\delta \xi} \tag{230}$$

where l is the reduced Lagrangian defined by

$$l : \mathfrak{g} \to \mathbb{R}, \quad l(\xi) = L(e, \xi) \tag{231}$$

The following theorem is a consequence of this construction:

Theorem 4.1 (Euler Poincaré Reduction)**.** *Let G be a Lie group and $L : TG \to \mathbb{R}$ a left-invariant Lagrangian. Define the reduced Lagrangian*

$$l : \mathfrak{g} \to \mathbb{R},\ l(\xi) := L(e, \xi) \tag{232}$$

as the restriction of L to $\mathfrak{g}$. For a curve $\mathbf{g}(t) \in G$, let

$$\xi(t) = \mathbf{g}(t)^{-1}\dot{\mathbf{g}}(t) := T_{\mathbf{g}(t)}L_{\mathbf{g}(t)^{-1}}\dot{\mathbf{g}}(t) \in \mathfrak{g} \tag{233}$$

Then, the following four statements are equivalent:

1. *Hamilton's Principle*

$$\delta \int_a^b L(\mathbf{g}(t), \dot{\mathbf{g}}(t))dt = 0 \tag{234}$$

holds for variations among paths with fixed endpoints.

2. *$\mathbf{g}(t)$ satisfies the Euler–Lagrange equations for Lagrangian L defined on G.*

3. *Hamilton's Principle*

$$\delta \int_a^b l(\xi(t))dt = 0 \tag{235}$$

holds on $\mathfrak{g}$, using variations of the form $\delta\xi = \dot{\eta} + [\xi, \eta]$, where $\eta(t)$ is an arbitrary path in $\mathfrak{g}$ that vanishes at the endpoints, i.e. $\eta(a) = \eta(b) = 0$.

4. *The (left invariant) Euler–Poincaré equations hold:*

$$\frac{d}{dt}\frac{\delta l}{\delta \xi} = -ad^*_\xi \frac{\delta l}{\delta \xi} \tag{236}$$

Example 4.1 (Heavy Symmetric Top – Lagrangian Formalism)**.** *The heavy top is a rigid body rotating with a fixed point of support (the 'pivot') in a constant gravitational field. Let m be the mass of the body, and let $\mathbf{k}$ be the vertical unit vector. Let χ be the vector from the point of support to the body's centre of mass. With this notation, the potential energy of the top is given by*

$$V(\mathbf{R}) = mg\langle \mathbf{k}, \mathbf{R}\chi \rangle \tag{237}$$

The kinetic energy is given by

$$K := \frac{1}{2}\int_{\mathcal{B}} \rho(X)||\dot{\mathbf{R}}X||^2 d^3X = \frac{1}{2}tr(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T) \tag{238}$$

We identify the Lagrangian $L : TSO(3) \to \mathbb{R}$ given by

$$L(\mathbf{R}, \dot{\mathbf{R}}) = \frac{1}{2}tr(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T) - mg\langle \mathbf{k}, \mathbf{R}\chi \rangle \tag{239}$$

While a free rigid body is $SO(3)$ left invariant, gravity breaks this symmetry. We will take advantage of this broken symmetry as follows: in order to regain $SO(3)$ left invariance, we define a new, extended Lagrangian which is defined over $SO(3) \times (\mathbb{R}^3)^*$, creating a degree of freedom related to the direction of the vertical axis. Let

$$L_{ext}(\mathbf{R}, \dot{\mathbf{R}}, \mathbf{v}, \dot{\mathbf{v}}) = \frac{1}{2}tr(\dot{\mathbf{R}}\mathbb{J}\dot{\mathbf{R}}^T) - mg\langle \mathbf{v}, \mathbf{R}\chi \rangle \tag{240}$$

Note that $L_{ext}(\mathbf{R}, \dot{\mathbf{R}}, \mathbf{v}, \dot{\mathbf{v}})\Big|_{\mathbf{v}=\mathbf{k}} = L(\mathbf{R}, \dot{\mathbf{R}})$.

It is easy to show that L_{ext} is left-invariant. That done, we write both angular velocity and $\mathbf{v}$ in body coordinates:

$$\hat{\mathbf{\Omega}} = \mathbf{R}^{-1}\dot{\mathbf{R}} \tag{241}$$

$$\mathbf{\Gamma} = \mathbf{R}^{-1}\mathbf{v} \tag{242}$$

Taking the time derivative, we see that

$$\dot{\mathbf{\Gamma}} := \frac{d}{dt}(\mathbf{R}^{-1}\mathbf{v}) = \dot{\mathbf{R}}^{-1}\mathbf{v} + \mathbf{R}^{-1}\dot{\mathbf{v}} \tag{243}$$

$$= -\mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{R}^{-1}\mathbf{v} + \mathbf{R}^{-1}\dot{\mathbf{v}} \tag{244}$$

$$= -\hat{\mathbf{\Omega}}\mathbf{\Gamma} + \mathbf{R}^{-1}\dot{\mathbf{v}} \tag{245}$$

$$= \mathbf{\Gamma} \times \mathbf{\Omega} + \mathbf{R}^{-1}\dot{\mathbf{v}} \tag{246}$$

With this in place we define $\tilde{L}_{ext}$ *to be the Lagrangian of the system with respect to its body coordinates, i.e.*

$$\tilde{L}_{ext}(\mathbf{R}, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}, \dot{\mathbf{\Gamma}}) = L_{ext}(\mathbf{R}, \mathbf{R}\hat{\mathbf{\Omega}}, \mathbf{R}\mathbf{\Gamma}, \mathbf{R}\dot{\mathbf{\Gamma}}) = L_{ext}(\mathbf{R}, \dot{\mathbf{R}}, \mathbf{v}, \dot{\mathbf{v}}) \tag{247}$$

Hamilton's Principle can now be applied to the new coordinates via

$$\delta \int_a^b \tilde{L}_{ext}(\mathbf{R}, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}, \dot{\mathbf{\Gamma}})dt = 0 \tag{248}$$

We now have to consider the variations $\delta\mathbf{R}$*,* $\delta\mathbf{\Omega}$*,* $\delta\mathbf{\Gamma}$*, and* $\delta\dot{\mathbf{\Gamma}}$*. However, since the heavy symmetric top is a rigid body,* $\tilde{L}_{ext}$ *is independent of* $\dot{\mathbf{\Gamma}}$*. Also note that* $\tilde{L}_{ext}$ *inherits left invariance from* L_{ext}*, prompting the translation*

$$\tilde{L}_{ext}(\mathbf{R}, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}) = \tilde{L}_{ext}(\mathbf{R}^{-1}\mathbf{R}, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}) = \tilde{L}_{ext}(\mathbb{I}_3, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}) \tag{249}$$

This implies that $\tilde{L}_{ext}$ *is independent of* $\mathbf{R}$ *as well. Therefore the only variations we need concern ourselves with are those of* $\mathbf{\Omega}$ *and* $\mathbf{\Gamma}$*. Starting with the former, we have*

$$\delta\hat{\mathbf{\Omega}} = -\mathbf{R}^{-1}\delta\mathbf{R}\mathbf{R}^{-1}\dot{\mathbf{R}} + \mathbf{R}^{-1}\delta\dot{\mathbf{R}} \tag{250}$$

$$= -(\mathbf{R}^{-1}\delta\mathbf{R})(\mathbf{R}^{-1}\dot{\mathbf{R}}) + \mathbf{R}^{-1}\delta\dot{\mathbf{R}} \tag{251}$$

$$= -(\mathbf{R}^{-1}\delta\mathbf{R})\hat{\mathbf{\Omega}} + \mathbf{R}^{-1}\delta\dot{\mathbf{R}} \tag{252}$$

Defining

$$\hat{\mathbf{\Sigma}} = \mathbf{R}^{-1}\delta\mathbf{R} \tag{253}$$

and computing

$$\frac{d\hat{\mathbf{\Sigma}}}{dt} = -\mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{R}^{-1}\delta\mathbf{R} + \mathbf{R}^{-1}\delta\dot{\mathbf{R}} \tag{254}$$

We can isolate for the last term in the sum to find

$$\mathbf{R}^{-1}\delta\dot{\mathbf{R}} = \frac{d\hat{\mathbf{\Sigma}}}{dt} + \mathbf{R}^{-1}\dot{\mathbf{R}}\hat{\mathbf{\Sigma}} = \frac{d\hat{\mathbf{\Sigma}}}{dt} + \hat{\mathbf{\Omega}}\hat{\mathbf{\Sigma}} \tag{255}$$

giving

$$\delta\hat{\mathbf{\Omega}} = \frac{d\hat{\mathbf{\Sigma}}}{dt} + [\hat{\mathbf{\Omega}}, \hat{\mathbf{\Sigma}}] \tag{256}$$

or in vector notation,

$$\delta\mathbf{\Omega} = \dot{\mathbf{\Sigma}} + \mathbf{\Omega} \times \mathbf{\Sigma} \tag{257}$$

Proceeding to the latter term, we find

$$\delta\mathbf{\Gamma} = \delta(\mathbf{R}^{-1}\mathbf{k}) = (\delta\mathbf{R}^{-1})\mathbf{k} \tag{258}$$

$$= -(\mathbf{R}^{-1}\delta\mathbf{R}\mathbf{R}^{-1})\mathbf{k} \tag{259}$$

$$= -(\mathbf{R}^{-1}\delta\mathbf{R})(\mathbf{R}^{-1}\mathbf{k}) \tag{260}$$

$$= -\hat{\mathbf{\Sigma}}\mathbf{\Gamma} = \Gamma \times \mathbf{\Sigma} \tag{261}$$

We now have a clear candidate for the reduced Lagrangian. Set

$$l(\mathbf{\Omega}, \mathbf{\Gamma}) := \tilde{L}_{ext}(\mathbb{I}_3, \hat{\mathbf{\Omega}}, \mathbf{\Gamma}, \dot{\mathbf{\Gamma}}) = \frac{1}{2} tr(\hat{\mathbf{\Omega}} \mathbb{J} \hat{\mathbf{\Omega}}^T) - mg\langle \mathbf{\Gamma}, \chi \rangle \tag{262}$$

or in vector notation for $\mathfrak{so}(3)$*,*

$$l(\mathbf{\Omega}, \mathbf{\Gamma}) := \tilde{L}_{ext}(\mathbb{I}_3, \mathbf{\Omega}, \mathbf{\Gamma}, \dot{\mathbf{\Gamma}}) = \frac{1}{2} tr(\mathbf{\Omega} \mathbb{I} \mathbf{\Omega}^T) - mg\langle \mathbf{\Gamma}, \chi \rangle \tag{263}$$

which, discarding $\delta\mathbf{R}$ *and* $\delta\dot{\mathbf{\Gamma}}$*, must satisfy*

$$\delta \int_a^b l(\mathbf{\Omega}, \mathbf{\Gamma}) dt = 0 \tag{264}$$

Now we derive the equations of motion for the heavy top. From the variational principle, we obtain

$$\begin{aligned}
0 &= \delta \int_a^b l(\mathbf{\Omega}, \mathbf{\Gamma}) dt && (265)\\
&= \int_a^b \langle \frac{\delta l}{\delta \mathbf{\Omega}}, \delta\mathbf{\Omega} \rangle dt + \int_a^b \langle \frac{\delta l}{\delta \mathbf{\Gamma}}, \delta\mathbf{\Gamma} \rangle dt && (266)\\
&= \int_a^b \langle \frac{\delta l}{\delta \mathbf{\Omega}}, \dot{\mathbf{\Sigma}} + \mathbf{\Omega} \times \mathbf{\Sigma} \rangle dt + \int_a^b \langle \frac{\delta l}{\delta \mathbf{\Gamma}}, \mathbf{\Gamma} \times \mathbf{\Sigma} \rangle dt && (267)\\
&= \int_a^b \langle -\frac{d}{dt} \left(\frac{\delta l}{\delta \mathbf{\Omega}} \right) + \frac{\delta l}{\delta \mathbf{\Omega}} \times \mathbf{\Omega} + \frac{\delta l}{\delta \mathbf{\Gamma}} \times \mathbf{\Gamma}, \mathbf{\Sigma} \rangle dt && (268)
\end{aligned}$$

$$\implies \frac{d}{dt} \left(\frac{\partial l}{\partial \mathbf{\Omega}} \right) = \frac{\partial l}{\partial \mathbf{\Omega}} \times \mathbf{\Omega} + \frac{\partial l}{\partial \mathbf{\Gamma}} \times \mathbf{\Gamma} \tag{269}$$

Since we have $\delta l / \delta \mathbf{\Omega} = \mathbb{I}\mathbf{\Omega}$ *and* $\delta l / \delta \mathbf{\Gamma} = mg\chi$*, the equations of motion are*

$$\begin{cases} \mathbb{I}\dot{\mathbf{\Omega}} = \mathbb{I}\mathbf{\Omega} \times \mathbf{\Omega} - mg\chi \times \mathbf{\Gamma} \\ \dot{\mathbf{\Gamma}} = \mathbf{\Gamma} \times \mathbf{\Omega} \end{cases} \tag{270}$$

4.3 Lie-Poisson Reduction

Now that we have defined the reduced Lagrangian, we can proceed with defining the reduced Hamiltonian. Our new Hamiltonian will be related to its corresponding Lagrangian in the exact same way as their unreduced counterparts: through a Legendre transformation.

Definition 4.1 (Reduced legendre transform)**.** *Let* $L : TG \to \mathbb{R}$ *be a hyperregular, left invariant Lagrangian. Let* $l : \mathfrak{g} \to \mathbb{R}$ *be the reduced Lagrangian such that* $l(\xi) = L(e, \xi)$*. We define the* ***reduced Legendre transform*** *as the map*

$$fl : \mathfrak{g} \to \mathfrak{g}^*, \ \langle fl(\xi), \eta \rangle := \frac{d}{ds}\Big|_{s=0} l(\xi + s\eta) = \langle \frac{\delta l}{\delta \xi}, \eta \rangle \tag{271}$$

for all $\xi, \eta \in \mathfrak{g}$*.*

Definition 4.2 (Reduced energy function, reduced Hamiltonian)**.** *We further define the* ***reduced energy function****:*

$$\tilde{e} : \mathfrak{g} \to \mathbb{R}, \ \tilde{e}(\xi) := \langle fl(\xi), \xi \rangle - l(\xi) \tag{272}$$

and the ***reduced Hamiltonian****:*

$$h : \mathfrak{g}^* \to \mathbb{R}, \ h(\mu) := \tilde{e} \circ fl^{-1} \tag{273}$$

Letting $\mu = fl(\xi)$, we have $h(\mu) = \tilde{e} \circ (fl)^{-1}(\mu)$. Identifying $fl(\xi)$ with $\delta l / \delta \xi$, we obtain the following:

Theorem 4.2 (Lie-Poisson Reduction)**.** *With μ and h as defined above, the Euler–Poincaré equations are equivalent to the Lie–Poisson equations*

$$\dot{\mu} = ad^*_{\partial h/\partial \mu}\mu \tag{274}$$

Example 4.2 (Heavy Symmetric Top – Hamiltonian Formalism)**.** *From the Euler-Poincaré reduction of the heavy symmetric top we found*

$$l(\mathbf{\Omega}, \mathbf{\Gamma}) = \frac{1}{2}\langle \mathbf{\Omega}, \mathbb{I}\mathbf{\Omega}\rangle - \langle mg\mathbf{\Gamma}, \chi\rangle \tag{275}$$

Applying the reduced Legendre transformation to $\mathbf{\Omega}$*, we obtain*

$$\mathbf{\Pi} := \frac{\delta l}{\delta \mathbf{\Omega}} = \mathbb{I}\mathbf{\Omega} \tag{276}$$

This corresponds to the reduced Hamiltonian

$$h(\mathbf{\Pi}, \mathbf{\Gamma}) = \frac{1}{2}\langle \mathbf{\Pi}, \mathbb{I}\mathbf{\Pi}\rangle + \langle mg\mathbf{\Gamma}, \chi\rangle \tag{277}$$

with equations of motion

$$\frac{d\mathbf{\Pi}}{dt} = \mathbf{\Pi} \times \mathbb{I}^{-1}\mathbf{\Omega} + mg\mathbf{\Gamma} \times \chi \tag{278}$$

$$\frac{d\mathbf{\Gamma}}{dt} = \mathbf{\Gamma} \times \mathbf{\Omega} \tag{279}$$

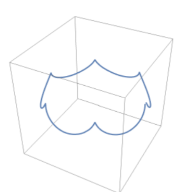
In searching for a physical interpretation of these equations of motion, we arrive at the following:

1. $\mathbf{\Pi} \times \mathbb{I}^{-1}\mathbf{\Omega}$ corresponds to a **centripetal** acceleration, scaling as second order in $\mathbf{\Omega}$.

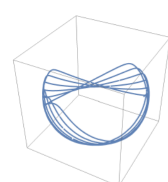
2. $mg\mathbf{\Gamma} \times \chi$ is a **precession** term, transferring the angular momentum of the top about its symmetry axis to a portion about the vertical axis. This causes the top to precess about its pivot point.

3. $\mathbf{\Gamma} \times \mathbf{\Omega}$ is a **nutation** term, accounting for the change in the angle the top makes with the vertical as it precesses.

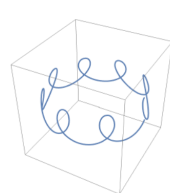
Taken in aggregate, these equations can be made tangible by visualizing the pattern such a top would "trace out" on the inside of a sphere centered at its pivot. To wrap up our treatment of he heavy symmetric top, we include three of these patterns here.



(a) Nutation of "cusp" type.



(b) Nutation of "smooth" type.



(c) Nutation of "loop" type.

Figure 5: Patterns traced out by the tip of a heavy symmetric top as it precesses about its pivot.

References

[1] Darryl D. Holm, Tanya Schmah, and Cristina Stoica. *Geometric Mechanics and Symmetry*. Oxford University Press, 2009.

[2] John Milnor. Analytic proofs of the “hairy ball theorem” and the brouwer fixed point theorem. *The American Mathematical Monthly*, 1978.

[3] Loring W. Tu. *An Introduction to Manifolds*. Springer, 2008.

Images

[4] Vedat Tanriverdi. Motion of a spinning symmetric top. Last accessed 07 June 2025.