Applied Mathematics in Robotics

Ragesh Kumar Ramachandran

Summer 2019

Contents

Abstract

These are the lecture notes prepared by the author for the seminar course offered in summer 2019 at the computer science department of University of Southern California. The notes covers important areas of mathematics with its emphasizes in robotics applications. The manuscript gives a brief overview on various mathematical subjects such as real analysis, abstract algebra, linear algebra, dynamical systems, topology and differential geometry. The primary goal of this seminar course is to introduce some important topics in mathematics and give a pictorial view that can be associated with these concepts. Therefore, the author do not assume any background for the reader other than high school mathematics. But, a good intuitive understanding of multivariate calculus is recommended.

1 Analysis

1.1 Why analysis?

Analysis is the rigorous study of such objects, with a focus on trying to pin down precisely and accurately the qualitative and quantitative behavior of these objects. *Complex analysis,* which concerns the analysis of the complex numbers and complex functions, *harmonic analysis*,which concerns the analysis of harmonics (waves) such as sine waves, and how they synthesize other functions via the Fourier transform, *functional analysis*, which focuses much more heavily on functions (and how they form things like vector spaces), and so forth.*Real analysis* is the theoretical foundation which underlies calculus, which is the collection of computational algorithms which one uses to manipulate functions [27]. The idea for real analysis is to identify the constitution of real numbers and function between real numbers. In general, understanding real analysis helps you answer the following questions:

1. What is a real number? Is there a largest real number? After 0, what is the "next" real number (i.e., what is the smallest positive real number)? Can you cut a real number into pieces infinitely many times? Why does a number such as 2 have a square root, while a number such as -2 does not? If there are infinitely many reals and infinitely many rationals, how come there are "more" real numbers than rational numbers?

2. How do you take the limit of a sequence of real numbers? Which sequences have limits and which ones don't? If you can stop a sequence from escaping to infinity, does this mean that it must eventually settle down and converge? Can you add infinitely many real numbers together and still get a finite real number? Can you add infinity many rational numbers together and end up with a non-rational number? If you rearrange the elements of an infinite sum, is the sum still the same?

3. What is a function? What does it mean for a function to be continuous? differentiable? integrable? bounded? can you add infinitely many functions together? What about taking limits of sequences of functions? Can you differentiate an infinite series of functions? What about integrating? If a function f(x) takes the value of f(0) = 3 when x = 0 and f(1) = 5 when x = 1, does it have to take every intermediate value between 3 and 5 when x goes between 0 and 1? Why?

It is a fair question to ask, "why bother?", when it comes to analysis. There is a certain philosophical satisfaction in knowing why things work, but a pragmatic person may argue that one only needs to know how things work to do real-life problems. The calculus training you receive in introductory classes is certainly adequate for you to begin solving many problems in physics, chemistry, biology, economics, computer science, finance, engineering, or whatever else you end up doing - and you can certainly use things like the chain rule, L'Hôpital's rule, or integration by parts without knowing why these rules work, or whether there are any exceptions to these rules. However, one can get into trouble if one applies rules without knowing where they came from and what the limits of their applicability are. Lets see some examples in which several of these familiar rules, if applied blindly without knowledge of the underlying analysis, can lead to disaster.

Example 1. *(Division by zeros). Consider this proof:*

$$\text{Let } x = y \neq 0 \tag{1.1}$$
$$x^2 = xy \tag{1.2}$$
$$x^2 - y^2 = xy - y^2 \tag{1.3}$$
$$(x+y)(x-y) = (x-y)y \tag{1.4}$$
$$x + y = y \tag{1.5}$$
$$2x = x \tag{1.6}$$
$$2 = 1 \tag{1.7}$$

Example 2. *(Divergent series). Consider the series,*

$$S_1 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \tag{1.8}$$

and the series,

$$S_2 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \tag{1.9}$$

S_1 is a slowly diverging series whereas S_2 is a converging series. One has to be careful while working with series since only operations with converging series yields meaningful results.

These notes merely touch upon some of the concepts in real analysis. Readers interested in diving deep into these topics are directed to the following references: [18, 22, 27].

1.2 Some definitions

Definition 1. A set is countable if it is finite or has the same cardinality as $\mathbb{N}$.
Example 3. *A, B, C, D, the set of natural number.*
Definition 2. A set is uncountable if it is not countable.
Example 4. *The set of real numbers, set of complex numbers*
Theorem 1.1. *The union of a countable collection of countable sets is countable.*

Is the set of rational numbers countable or uncountable? ans: countable.
Definition 3. (Rational number construction). The equivalence classes of integer pairs such at pairs (a, b) and (c, d) are equivalent if $a \cdot d = c \cdot b$. Usually denotes as $\mathbb{Q}$.
Definition 4. A sequence is a function from $\mathbb{N}$ into some set.
Definition 5. A x_n is a subsequence of y_n if there exists integers $1 < k_1 < k_2 < k_3 \cdots$ such that:

$$x_n = y_{k_n}$$

Theorem 1.2. *Let E be the set of all sequences whose terms are the digits 0 and 1.Then, E is uncountable*

An interesting consequence of the above is that, almost every decision problem is unsolvable by any program. The way we prove this statement is by associating every program to a binary natural number and showing the space of all decision problem is isomorphic(equivalent) to $\mathbb{R}$. Any decision problem is a map from at least a countably infinite set to $\{0, 1\}$. Therefore, a decision problem is a infinite sequence of 0 and 1, and the set of decision problems is uncountable. But the space of programs has the same cardinality as $\mathbb{N}$.

Definition 6. Let $\mathbb{A} \subset \bar{\mathbb{R}}$, then b is a real number is an *upper bound* for $\mathbb{A}$ if $\mathbb{A} \subset [-\infty b]$. Similarly, one can define a *lower bound*. The least upper bound is called the *supremum* of $\mathbb{A}$ and greatest lower bound is called the *infimum*.
Definition 7. (Limits). We can also define limits of sequence using supremum and infimum of increasing and decreasing respectively. For a decreasing sequence x_n, $\lim x_n$ is $\inf x_n$ and similarly for an increasing sequence.

Let,

$$\vec{x}_m = \inf_{n \geq m} x_n, \quad \bar{x}_m = \sup_{n \leq m} x_n, \ m \in \mathbb{N} \tag{1.10}$$

Definition 8. (Limit inferior).

$$\liminf x_n = \lim \vec{x}_n = \sup_m \inf_{n \geq m} x_n. \tag{1.11}$$

Definition 9. (Limit superior).

$$\limsup x_n = \lim \bar{x}_n = \inf_m \sup_{n \geq m} x_n. \tag{1.12}$$

1.3 Metric Spaces

Definition 10. A *space* is a set with some additional structure.

A metric space is a tuple $(\mathbb{E}, d)$, where $\mathbb{E}$ is a set and $d : \mathbb{E} \times \mathbb{E} \to \mathbb{R}^+$ is metric function. A metric function satisfies the following condition $\forall x, y, z \in \mathbb{E}$:

1. $d(x, y) = d(y, x)$.
2. $d(x, y) = 0$ if and if only if $x = y$.
3. $d(x, y) + d(y, z) \geq d(x, z)$.

Example 5. *Euclidean spaces. Metric space* $(\mathbb{R}^n, \|\cdot\|_2)$.
Example 6. *The space of continuous function in the interval* $[0, 1]$. *The distance function for the space:*

$$d(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|. \tag{1.13}$$

Definition 11. Distance to a point from a set. Let $(\mathbb{E}, d)$ be a metric space. If $x \in \mathbb{E}$ and $A \subset \mathbb{E}$, then

$$d(x, A) = inf\{d(x, y) : y \in A\} \tag{1.14}$$

Definition 12. Diameter of a set. The diameter of a set $A \subset \mathbb{E}$ is defined as:

$$diam A = sup\{d(x, y) : x \in A, y \in A\}. \tag{1.15}$$

Definition 13. Open ball defined on a metric space $(\mathbb{E}, d)$ centered at x with radius r:

$$B(x, r) = \{y \in \mathbb{E} : d(x, y) < r\}. \tag{1.16}$$

Definition 14. A set A is said to be *open* if for every $x \in A$ there is an $r > 0$ such that $B(x, r) \subset A$. A set is said to be *closed* if it complement is open.

Example 7. *If the metric space is* $\mathbb{R}$*, then* $(a,b),(-\infty,b),(a,\infty)$ *are open sets and* $[a,b]$ *is a closed set.*

Properties:

1. Arbitrary union of open sets are open.

2. Finite intersections of open sets are open.

Definition 15. Convergence of sequence. A sequence x_n in $(\mathbb{E},d)$ is said to be convergent, if there exist an element $x \in \mathbb{E}$, such that $\lim d(x,x_n) = 0$.

The definition includes implicitly the fact that the limit is unique. It is easy to show this through triangle inequality. There is an interesting theorem about convergent sequences and closed set. If we think of particle in a closed set jumping from one element in a convergent sequence to another. The particle can never escape the closed set.

Theorem 1.3. *A set is closed if and only if it contains the limits of every convergent sequence in it.*

The definition of convergence is not very useful in practice. If you think about it, the definition requires knowledge of the sequence limit which if you know in advance then problem is solved in the first place.

Definition 16. (Cauchy sequence). A sequence is a *Cauchy* sequence if for every $\epsilon > 0$ there exist a n_ϵ such that $d(x_m,x_n) < \epsilon$ for every $m > n \geq n_\epsilon$.

Cauchy sequences are bound. Convergent sequence are Cauchy. But the reverse is not always true.

Example 8. *Consider the following sequence of rational numbers in* $\mathbb{Q}$*,*

$$q_n = (1+\frac{1}{n})^n.$$

It is easy to show that the above sequence is a Cauchy sequence. But the sequence converges to e which is an irrational number. Therefore, the Cauchy sequence does not converge to a point in set of rational numbers.

Definition 17. (Complete Metric). A metric space is complete if every Cauchy sequence in the space is a convergent sequence.

1.3.1 Construction of real number from Cauchy sequence

One natural way to construct complete space from and an existing metric space is by taking every possible Cauchy sequence limit in the metric space and adding it to space yielding a new complete metric space. Interestingly, If we take $\mathbb{Q}$ and add the limits of all its Cauchy sequence we get the real line. This is one way to construct real number. Euclidean space is a complete metric space.

1.4 Continuous maps or functions between metric spaces

A map or function f, $f : \mathbb{A} \longrightarrow \mathbb{B}$, connects each element of a set $\mathbb{A}$ (domain set) to an element of a set $\mathbb{B}$ (target set).

Definition 18. A map $f : \mathbb{A} \longrightarrow \mathbb{B}$ is locally *continuous* at a point $x \in \mathbb{A}$ if there exists an open set $\mathbb{V} \subset \mathbb{B}$ containing $f(x)$ such that $f^{-1}(\mathbb{V})$(preimage) is an open set of $\mathbb{A}$.

This definition can be extended to talk about continuity of a map in a global sense:

Definition 19. A map $f : \mathbb{A} \longrightarrow \mathbb{B}$ is continuous if preimages of all open sets in $\mathbb{B}$ are open set of $\mathbb{A}$.

1.5 Compactness

Given an metric space with an underlying set $\mathbb{E}$. We define the following.
Definition 20. (Open cover). A collection of open sets $\{\mathbb{C}_i \in \mathbb{E}\}_{i\in\mathbb{I}}$ is said to be an *open cover* of $\mathbb{A} \subset \mathbb{E}$, if $\mathbb{A} \subset \underset{i\in\mathbb{I}}{\cup}\mathbb{C}_i$.
Definition 21. (Compact set). A set $\mathbb{A} \subset \mathbb{E}$ is said to be a *compact* if for its every open cover there exist a finite sub-cover containing the set.

Compactness gives an intuition that at the set is not "huge" and therefore is easier to work with. Most mathematical theorems are first studied in a compact setting before being generalized to much "larger" spaces.

Some properties of compact sets in metric spaces:

1. Every compact set is bounded
2. Every closed subset of a compact set is compact
3. Every compact set is closed
4. Every compact set is complete

Every compact set is closed and bounded. The converse is not true in general. The next theorem states that it is true for Euclidean spaces.
Theorem 1.4. *(Heine-Borel Theorem). Every subset of an Euclidean space is compact if and only if it is closed and bounded.*
Theorem 1.5. *Image of a continuous map from a compact set is compact.*

Why do we care? This property is the key to existence of global optima in optimization problem. A cost function is a mapping from our search space of interest to $\mathbb{R}$. If our search space is compact then by using Theorem 1.5 we can show the cost function attains its maxima and minma in the search space. This idea is key to the famous *Weierstrass Extreme Value Theorem.*

1.6 Dense sets

Definition 22. (A dense set). A set $\mathbb{A} \subset \mathbb{E}$ is a *dense* set if for every open set containing $e \in \mathbb{E}$ contains at least one point from $\mathbb{A}$.
Example 9. *Rational and irrational numbers are dense set of real numbers. Moreover, rational number set is countable dense set. Note that rational numbers and irrational number are complements to each other.*

The concept of dense set are very important for approximation theorems. Whenever we are trying to find a good approximation to some set, we are ideally looking for a dense subset of the set that can be handled easily. In an important theorem useful in practice concerning dense sets is the *Weierstrass approximation theorem*, which states that every continuous function on a closed interval can be approximated as closely as desired by a polynomial function.

1.7 Integration

Definition 23. Let I be a nonempty, compact interval. A *partition* of I is a finite collection $\mathbf{P} = \{I_1, I_2, I_3, \cdots I_n\}$ disjoint interval such that their union is I. If $I = [a, b]$ and $a = x_0 < x_1 < \cdot < x_{n-1} < x_n = b$, then $I_k = [x_{k-1}, x_k]$.

Definition 24. Let f be a bounded function defined on $[a,b]$, then $M_k = \sup\{f(t) : t \in I_k\}$ and $m_k = \inf\{f(t) : t \in I_k\}$. We define the upper sum associated with the partition $\mathbf{P}$ as $U(f,\mathbb{P}) = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$. Similarly, the lower sum associated with the partition $\mathbf{P}$ as $L(f,\mathbb{P}) = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$.

Clearly, $L(f,\mathbb{P}) \leq \int_{[0,1]} f dx \leq U(f,\mathbb{P})$.

Definition 25. (Riemann integral). $\sup_{\mathbb{P}} L(f,\mathbb{P}) = \int_{[0,1]} f dx = \inf_{\mathbb{P}} U(f,\mathbb{P})$

The Riemann integral has certain issues for example consider the function $\mathbb{1}_{\mathbb{Q}}$ defined as:

$$\mathbb{1}_{\mathbb{Q}}(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{otherwise} \end{cases} \tag{1.17}$$

We cannot compute the Riemann integral of this function. Why?

This is where a new kind of integral called the *Lebesgue integral* comes for help. The idea is instead of partitioning the domain of the function we partition the codomain of the function to compute the integral.

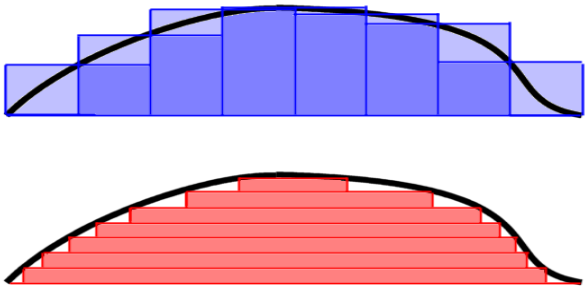


Figure 1.1: A figure illustrating the idea of Riemann integral(top) and Lebesgue integral(botton)(Image from Wikipedia). (Top) We construct vertical strips to compute a Riemann integral. (Bottom) We construct horizontal strip to compute a Lebesgue integral.

To define the Lebesgue integral requires the formal notion of a measure that, roughly, associates to each subset $\mathbb{A}$ of real numbers a nonnegative number $\mu(\mathbb{A})$ representing the "size" of $\mathbb{A}$. This notion of "size" should agree with the usual length of an interval or disjoint union of intervals. Suppose $y = f(x)$ is a bounded non negative function defined on $[a,b]$, then area of an horizontal strip between $y = t$ and $y = t + dt$ is given by $\mu(x : t < f(x) < t + dt)dt$(think of $\mu(\mathbb{A})$ as the area of the region containing the set $\mathbb{A}$). The total area under f can be computed by summing all the horizontal strips as:

$$\int_0^{\sup_{[a,b]} f(x)} \mu(x : t < f(x) < t + dt)dt \tag{1.18}$$

This is roughly the idea of a Lebesgue integral. A more common and elegant approach for constructing Lebesgue integral is achieved using the concept of *simple functions*

2 Coordinate free Linear Algebra

2.1 Motivation

Objects in real world exist independent of the coordinate system we use to describe them. For example, your school and house are place at various locations and, the distance between them is independent of the unit you choose to measure them. Therefore, it is important to describe these objects in a coordinate free way, so that our description of these object do not change with specific units we use to quantify them. For example, if we are developing a robotics algorithm for localization problem, it should not depend on the fact that we choose to align the robot's motion along one of the axis instead of another or the fact that we choose a polar coordinates instead of cartesian coordinates. In this section, we will discuss about how vector spaces can be viewed in a coordinate free manner. Of course, for computational purpose we will have to rely on a coordinate system to describe the object of interest. But it is important to keep mind about the geometry of the object that we have assigned a coordinate system to. We will begin this session by introducing the concepts from abstract algebra such as groups, rings, fields, and vector spaces. Thereafter, we will be focusing on vector spaces and maps between them. We will be return to ideas in group theory in a later session when we will be discussing about Lie groups. Some good references to the ideas presented in this session are [1, 9, 11].

2.2 Groups

Definition 26. (Group). A *group* $(\mathbb{G}, \circ)$ is a set $\mathbb{G}$ with the composition law $\circ : \mathbb{G} \times \mathbb{G} \longrightarrow \mathbb{G}$ which satisfies the following conditions or axioms:

1. (Associativity) $a \circ (b \circ c) = (a \circ b) \circ c$, $\forall a, b, c \in \mathbb{G}$.
2. (Identity element) There exist an element $e \in \mathbb{G}$ such that $a \circ e = e \circ a = a$.
3. (Inverse element) For every element $a \in \mathbb{G}$ there is an element denoted as a^{-1} such that $a \circ a^{-1} = a^{-1} \circ a = e$.

Note that, the group composition law need not be *commutative*. But if the group composition law is commutative then the group is term as an *abelian* group.

Examples of groups are: 1) Integers with the addition operation, 2) $\mathbb{R} - \{0\}$ with multiplication operation, 3) set of invertible $\mathbb{R}^{n \times n}$ matrices with the matrix multiplication operation 4) equivalence classes of integer mod n under addition.

Identify the abelian and non abelian groups in above example.

A group is **finite** if it underlying set contains finite elements. Likewise, a group has infinite order it the underlying set is infinite.

Some properties of groups:

1. The identity element in a group is unique.
2. Every element has an unique inverse.
3. $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$ for all $a, b \in \mathbb{G}$.
4. $(a^{-1})^{-1} = a$.
5. All exponent law hold for groups

Definition 27. (Subgroups). A set of group which is also a group when the group operation is restricted to the subset.
Example 10. *SO(2), SE(2) subgroups of invertible matrices, integers with addition operation.*

What are the conditions required for a subset of a group to be a subgroup.

1. The identity element of the original group should be contained in the subset.
2. associativity and existence of inverse elements.

2.3 Rings

Group are objects with a single binary operation satisfying some axiomatic conditions. But often we are interested in objects which allow two binary operations. For example, integers with the natural addition and multiplication operations. If one recalls the properties of these operation from his/her primary school, one could find that these two operations are related through something called as a distributive property.
Definition 28. An non empty set $(\mathbb{R})$ is a ring if it admits two binary operations $+$ and $\cdot$ satisfying the following conditions $\forall a, b, c \in (\mathbb{R})$:

1. $a + b = b + a$.
2. $(a + b) + c = a + (b + c)$.
3. There exists and element $0 \in (\mathbb{R})$ such that $a + 0 = 0 + a = a$.
4. For every a there exist an unique element $-a \in (\mathbb{R})$ such that $a + -a = 0$.
5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
6. $(a + b) \cdot c = a \cdot c + b \cdot c$
   $a \cdot (b + c) = a \cdot b + a \cdot c$.

Example 11. $\mathbb{Q}$ *with natural addition and multiplication, continuous real values function on a closed interval*$(x^2, \sin(x))$. 3×3 *real matrices under matrix addition and multiplication.*

Note that a ring does not require to have an inverse element with respect to $\cdot$ operation. If we include this requirement, along with the commutative property of $\cdot$ into the definition of rings we end up with a new algebraic structure called *field*.

2.4 Field

Definition 29. A field is a set $\mathbb{F}$ together with two binary operations $+$ and $\circ$ satisfying the following conditions $\forall a, b, c \in \mathbb{F}$:

1. $a + b = b + a$.
2. $(a + b) + c = a + (b + c)$.
3. There exists and element $0 \in (\mathbb{R})$ such that $a + 0 = 0 + a = a$.
4. For every a there exist an unique element $-a \in \mathbb{F}$ such that $a + -a = 0$.
5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
6. $a \cdot b = b \cdot a$.

7. $(a+b)\cdot c = a\cdot c + b\cdot c$
   $a\cdot(b+c) = a\cdot b + a\cdot c.$
8. There exists and element $1 \in \mathbb{F}$ such that $a\cdot 1 = 1\cdot a = a$.
9. For every $a \neq 0$ there exist an unique element $a^{-1} \in \mathbb{F}$

Fields are very common in mathematical constructs used in algebra. The various examples of fields include: 1) real number with usual addition and multiplication, 2) Complex numbers.

2.5 Vector space

Definition 30. A *vector space* $\mathbb{V}$ over a field $\mathbb{F}$ is an abelian group with scalar product $\alpha \odot V$ or αV for all $\alpha \in \mathbb{F}$ and all $V \in \mathbb{V}$ satisfying the following axiom:

1. $\alpha(\beta V) = (\alpha\beta)V$;
2. $(\alpha+\beta)V = \alpha V + \beta V$;
3. $\alpha(V+U) = \alpha V + \alpha U$;
4. $1V = V$;

for all $\alpha, \beta \in \mathbb{F}$ and $V, U \in \mathbb{V}$

An element in a vector space is called a *vector*.
Example 12. *If $\mathbb{F}$ is a field, then $F[x]$ is a vector space over $\mathbb{F}$. The vectors in $F[x]$ are polynomials.*
Example 13. *The set of all continuous real-valued functions on a closed interval is a vector space over $\mathbb{R}$. This example can be generalized to set of all continuous functions on a closed set to any field $\mathbb{F}$.*
Definition 31. $\mathbb{V} = \{a + b\sqrt{3} : a, b \in \mathbb{Q}\}$ is a vector space over $\mathbb{Q}$.
Example 14. *The set $\mathbb{F}^\infty$ is the set of all infinite sequences of elements of $\mathbb{F}$.*

In short, a vector space is a set of objects that can be added together and scaled in nice ways. A vector space is a real vector space if the underlying field is $\mathbb{R}$. Similarly, a vector space is termed as a complex vector space if $\mathbb{F}$ is the space of complex numbers.

2.6 Subspace

A subset $\mathbb{U} \subset \mathbb{V}$ is subspace of $\mathbb{V}$ if $\mathbb{U}$ is a vector space.
Example 15. *The line y=x is a subspace in $\mathbb{R}^2$.*
Example 16. *The set of all continuous functions defined on the closed interval $[0,1]$ is a subspace of $\mathbb{R}^{[0,1]}$(set of all function in the interval $[0,1]$).*
Example 17. *The set of differentiable functions on the open interval $(0,4)$ such that $f'(2) = b$ is a subspace of $\mathbb{R}^{(0,4)}$ if and only if $b = 0$. Why?*
Because $f(x) = 0$ should be in the subspace which implies $f'(x) = 0$ should be present in the subspace.

2.7 Span, Linear independence and Basis

A linear combination of the elements in set $\{V_1, V_2, \cdots V_m\}$ is defined as: $\sum_{i=1}^{m} \alpha_i V_i,\ \alpha_i \in \mathbb{F}$.

The span of $\{V_1, V_2, \cdots V_m\}$ is the set of all linear combination.

The elements in a set $\{V_1, V_2, \cdots V_m\}$ are linearly independent if and only if no element in the set can be written as a linear combination of the other elements.

A vector space is finite dimensional if the entire space can be spanned by a finite number elements. A vector space is infinite dimensional if it is not finite dimensional.

A basis of a vector space is a collection of elements in the vector space which are linearly independent and span the entire space.

2.8 Linear maps

The vector space of all polynomials with coefficients in $\mathbb{F}$ is denoted by $\mathcal{P}(\mathbb{F})$.
Definition 32. A *linear map* from $\mathbb{V}$ to $\mathbb{W}$ is a map $\mathbf{T} : \mathbb{V} \longrightarrow \mathbb{W}$ such that: $\mathbf{T}(\alpha V_1 + \beta V_2) = \alpha \mathbf{T}(V_1) + \beta \mathbf{T}(V_2)$ $\alpha, \beta \in \mathbb{F},\ V_1, V_2 \in \mathbb{V}$.

The set of all linear map from $\mathbb{V}$ to $\mathbb{W}$ is denoted $\mathbb{L}(\mathbb{V}, \mathbb{W})$. This space is a vector space.
Example 18. $\mathbf{0} \in \mathbb{L}(\mathbb{V}, \mathbb{W})$, $\mathbf{I} \in \mathbb{L}(\mathbb{V}, \mathbb{W})$, *(Differentiation operation)* $\mathbf{D} \in \mathbb{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ *and (Integration)* $\mathbf{T}_I \in \mathbb{L}(\mathcal{P}(\mathbb{R}), \mathbb{R})$, $\mathbf{T}_I(p) = \int_0^1 p(x)dx$

2.8.1 Matrix representation of linear map

Suppose $\mathbf{T} \in \mathbb{L}(\mathbb{V}, \mathbb{W})$ and $\{V_1, V_2, \cdots V_n\}$ are a basis for $\mathbb{V}$ and $\{W_1, W_2, \cdots W_m\}$ are a basis for $\mathbb{W}$. The $m - n$ matrix with respect to these bases $\mathcal{M}(\mathbf{T})$ whose entities $A_{j,k}$ are defined by:

$$\mathbf{T}V_k = A_{1,k}W_1 + A_{2,k}W_2 + \cdots + A_{m,k}W_m$$

Note that the matrix representing the linear map depends on basis used for the spaces.
Example 19. *Consider the linear map* $\mathbf{T} \in \mathbb{L}(\mathbb{R}^2, \mathbb{R}^3)$*:*

$$T(x, y) = (3x + y, x + 4y, 2x + 6y)$$

$T(0, 1) = (1, 4, 6)$ *and* $T(1, 0) = (3, 1, 2)$*, then the matrix associated with the map with the standard basis is*

$$\mathcal{M}(\mathbf{T}) = \begin{pmatrix} 1 & 3 \\ 4 & 1 \\ 6 & 2 \end{pmatrix}$$

2.9 Invertibility and Isomorphism

Definition 33. A linear map $\mathbf{T} \in \mathbb{L}(\mathbb{V}, \mathbb{W})$ is invertible if there exist a linear map $\mathbf{S} \in \mathbb{L}(\mathbb{W}, \mathbb{V})$ such that $\mathbf{ST}$ equals the identity map on $\mathbb{V}$ and $\mathbf{TS}$ equals the identity map on $\mathbb{W}$.
Definition 34. A linear map $\mathbf{T} \in \mathbb{L}(\mathbb{W}, \mathbb{V})$ is the inverse of $\mathbf{T} \in \mathbb{L}(\mathbb{V}, \mathbb{W})$ if $\mathbf{ST} = \mathbf{I}$ and $\mathbf{TS} = \mathbf{I}$
Definition 35. Two vector spaces are isomorphic is there exist a invertible linear map between them.

Think isomorphism as relabeling of the elements.
Theorem 2.1. *Two finite dimension vector spaces are isomorphic over* $\mathbb{F}$ *if and only if they have the same dimensions*

2.10 Dual vectors(Covectors)

How can we multiply vectors?

Dual vectors gives a framework to do this. Strictly speaking this intuition is incorrect, as dual vector can be thought as an multiplication operator only in an inner product space specifically Hilbert space.

As a motivation, consider the following example. Lets consider the space of vegetables: onion, tomato and carrot. Now any purchase of these vegetable would be a point in this vector space: $3\vec{onion}+4\vec{tomato}+5\vec{carrot}$. Now lets say we want to compare two such purchases, for example $3\vec{onion}+4\vec{tomato}+5\vec{carrot}$ and $2\vec{onion}+1\vec{tomato}+15\vec{carrot}$. Since a notion of length, angle between vectors or a metric cannot defined for this vector space(no sense in that) it is not possible to perform this task. But if we define a linear price function p such that $p(3\vec{onion}+4\vec{tomato}+5\vec{carrot}) = 3p(\vec{onion}) + 4p(\vec{tomato}) + 5p(\vec{carrot})$, where $p(\vec{onion})$, $p(\vec{tomato})$ and $p(\vec{carrot})$ denote the unit price of the three vegetable respectively then we compare them on the basis of their total price. The space of price function is the dual space of vegetable vector space. Note that the price vector and vegetable vector are interchangeable.
Definition 36. A linear function on $\mathbb{V}$ is a linear map from $\mathbb{V}$ to $\mathbb{F}$ denoted as $\mathbb{L}(\mathbb{V}, \mathbb{F})$. This is denoted as $\mathbb{V}'$

In essence, a covector eats a vector and splits out an element in $\mathbb{F}$.
Example 20. $\phi(x, yz) : (3x + 8y + 5z)$, $\phi : \mathcal{P}(\mathbb{R}) \longrightarrow \mathbb{R}$, $\phi(p) = p''(3) + 4p(2)$, $\phi : \mathcal{P}(\mathbb{R}) \longrightarrow \mathbb{R}$, $\phi(p) = \int_0^1 p(x)dx$.
Example 21. *The function* $< V_1, \cdot >$ *where* $< \cdot, \cdot >$ *is the inner product in* $\mathbb{R}^n$ *and* $V_1 \in \mathbb{R}^n$.
Theorem 2.2. *For finite dimensional vector spaces the space and its dual has the same dimension.*

In finite dimensional vector spaces vectors can thought as column vectors and covectors can be thought as row vectors.
Definition 37. (Dual basis). Suppose $\{V_1, V_2 \cdots V_n\}$ is a basis for a vector space $\mathbb{V}$, then $\{\phi_1, \phi_2, \cdots \phi_n\}$ is a dual basis of $\mathbb{V}'$ such that:

$$\phi_j(v_k) = \begin{cases} 1, & \text{if } k = j \\ 0, & \text{otherwise} \end{cases}$$

The level set of an element of dual space form a family of parallel hyperplanes in vector space. In 2D this becomes parallel lines.

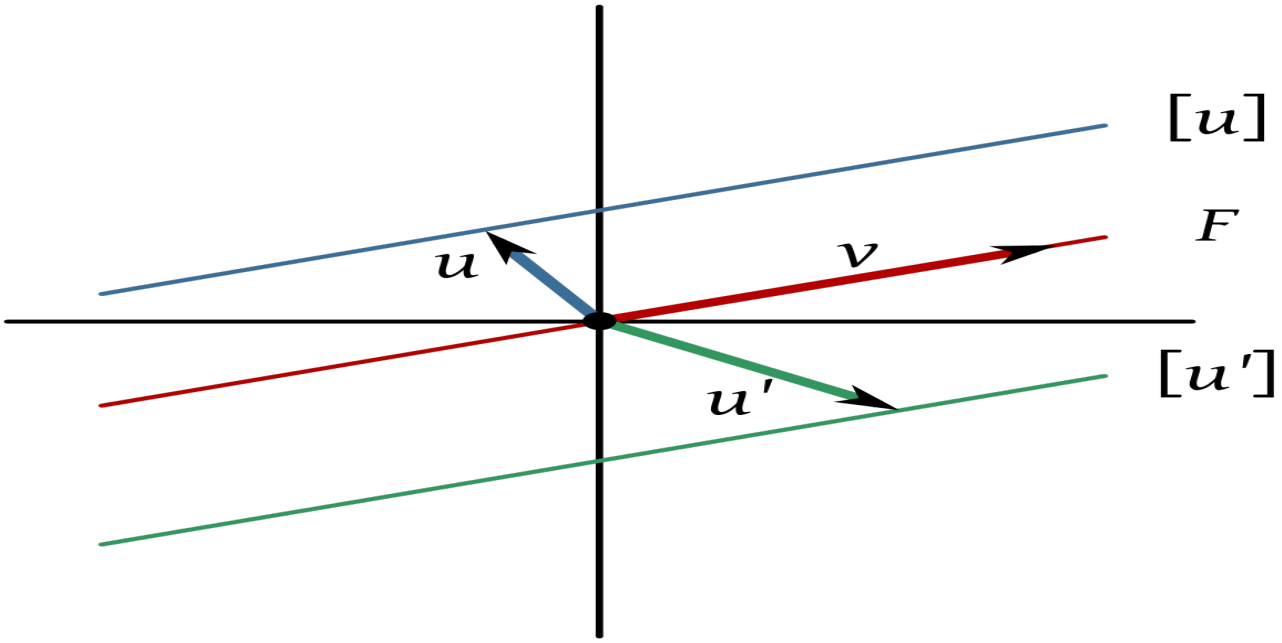
The relationship in the previous definition is called the *principle of duality*. The different choice of vector basis would yield different set of coordinate functions(dual vector basis), but principle of duality still remain the same.

2.11 Quotient space

Till now we looked at how to add vectors, when and how to multiply vectors. Now lets think about how to define division operation in a vector space. Think about how we divide 12 by 3, the answer is of course 4. The important thing to understand here is what we are doing in terms of the object involved in the operation. In other words, when we divide 12 by 3, what we are essentially asking is: if we had 12 objects(e.g. pencils) how many groups of 3 does the set of 12 objects contain. This intuition is used to define the notion of a quotient space.
Definition 38. (Cosets). For each element $V \in \mathbb{V}$, the *coset* of V with respect to the subspace $\mathbb{W}$ denotes by $V + \mathbb{W}$ is defined as:

$$V + \mathbb{W} = \{V + W : W \in \mathbb{W}\} \tag{2.1}$$

Theorem 2.3. *Let* $V_1, V_2 \in \mathbb{V}$, *then* $V_1 + \mathbb{W} = V_2 + \mathbb{W}$ *if and only if* $V_1 - V_2 \in \mathbb{W}$



If we define an equivalence relation such that V_1 V_2 if $V_1 - V_2 \in \mathbb{W}$, then we obtain an collections of equivalence classes each for every coset. The equivalence class of a coset $V + \mathbb{W}$ is commonly denoted as $[V]$.

Definition 39. (Quotient space). The *quotient space* of $\mathbb{V}$ with respect the subspace $\mathbb{W}$ is the set of all equivalence classes denoted as $\mathbb{V}/\mathbb{W}$.

Example 22. *Let* $\mathbb{W} = \{(x, 2x) : x \in \mathbb{R}\}$*, then* $\mathbb{R}^2/\mathbb{W}$ *is the set of lines in* $\mathbb{R}^2$ *parallel to* $\mathbb{W}$*.*

Example 23. *If* $\mathbb{W} = \{(x, y, 0) : x, y \in \mathbb{R}\}$*, then* $\mathbb{R}^3/\mathbb{W}$ *is the set of planes in* $\mathbb{R}^3$ *parallel to* $\mathbb{W}$*.*

3 Linear systems

3.1 Motivation

Researchers have been trying to analyze and understand the characteristics of general dynamical systems for eons. In early days, people believed that they could unravel the mysteries of nature and understand the way in which universe works in an absolute sense. Soon they realized that, the analysis of a system is much easier if one restricts the focus to specific characterizes of a system relevant to a problem of interest. This idea is generally referred as modeling of system. Linear models are a kind of model which have been well studied and deeply understood by researchers. Unfortunately, most reasonable models of systems are non-linear. Under certain conditions, one could approximate a non-linear system with a linear system(linearization) and understand the behavior of the non-linear system by analyzing the derived linear system. Non-linear system theory is still an active area of research and there are other ways to analyze non linear systems apart from linearizing them. In this section, we will concentrate on linear time invariant systems and discuss about various techniques to analyze them. In the successive section, we will look at some non-linear systems and explore some ways to understand the behavior of certain classes of them. Some good references to the ideas presented in this session are [7, 10, 12]. Readers interested in non-linear system theory and control may use the following references [13, 24, 29]. A more mathematically rigorous treatment of non-linear system and control theory can be found in [23, 25].

3.2 State space representation of linear systems

Consider following set of equations:

$$\dot{X}(t) = \mathbf{A}(t)X(t) + \mathbf{B}(t)U(t), \quad X(0) = X_0 \tag{3.1}$$

$$Y(t) = \mathbf{C}(t)X(t) + \mathbf{D}(t)U(t) \tag{3.2}$$

The above set of equations can be used to represent a continuous time linear system. These set of equations is referred to as a state space representation of a linear system. The variables used in the equations represent the following quantities of the underlying system:

- $t \in \mathbb{R}_{\geq 0}$: time variable
- $X(t) \in \mathbb{R}^n$: vector of n state variables
- $U(t) \in \mathbb{R}^m$: vector of m input variables
- $Y(t) \in \mathbb{R}^p$: vector of p output variables
- $X_0 \in \mathbb{R}^n$: initial condition of the system
- $\mathbf{A} \in \mathbb{R}^{n \times n}$: state matrix
- $\mathbf{B} \in \mathbb{R}^{n \times m}$: input matrix
- $\mathbf{C} \in \mathbb{R}^{p \times n}$: output matrix
- $\mathbf{D} \in \mathbb{R}^{p \times m}$: feedforward matrix.

When the matrices in the state space representation of a linear system are independent of time, then we term the system as a linear time invariant system. The first equation describes the evolution of the state with respect to time and is generally referred as *state equation*, and the next equation called *output equation* maps to state of the

system to some observable outputs of the system. If the linear time invariant system of interest evolves in discrete time then the dynamics of the system can be represented by the following discrete time state space equation.

$$X(k+1) = \mathbf{A}X(k) + \mathbf{B}U(k), \quad X(0) = X_0 \tag{3.3}$$

$$Y(k) = \mathbf{C}X(k) + \mathbf{D}U(k), \quad k = 0, 1, 2 \cdots \tag{3.4}$$

In this section, we will restrict our attention to discrete time state space system as they more intuitive and easier to understand, but keep in mind that the results transfer easily to continuous time linear systems in appropriate sense.

Example 24. *(Discrete time particle dynamics). Consider the following discrete time dynamics of a particle.*

$$P(k+1) = P(k) + TV(k) + \frac{T^2}{2}a(k) \tag{3.5}$$

$$V(k+1) = V(k) + Ta(k). \tag{3.6}$$

The state space form of the above dynamics can be written as:

$$\begin{bmatrix} P(k+1) \\ V(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P(k) \\ V(k) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix} a(k) \tag{3.7}$$

Now if the particle is equipped with a localization device (e.g. GPS) then the output equation of the system takes the form

$$P(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} P(k) \\ V(k) \end{bmatrix} \tag{3.8}$$

3.3 General state transfer

Given a linear time invariant discrete time dynamical system and sequence of l inputs $\{U(0), U(1), U(2), \cdots, U(l-1)\}$ then the system can be transferred from an initial state X_0 to some final state $X(l)$ if,

$$X(l) = \mathbf{A}^l X_0 + \sum_{\tau=0}^{l-1} \mathbf{A}^{l-1-\tau}\mathbf{B}U(\tau). \tag{3.9}$$

Similarly, for a linear time invariant continuous time system, the general state transfer equation given an input $U : [0, t] \longrightarrow \mathbb{R}^m$ can be expressed as,

$$X(t) = \exp^{\mathbf{A}t} X_0 + \int_0^t \exp^{\mathbf{A}(t-\tau)} \mathbf{B}U(\tau)d\tau. \tag{3.10}$$

3.4 Stability of autonomous system

A system of the form

$$X(k+1) = \mathbf{A}X(k) \tag{3.11}$$

is called an *autonomous system*. In other words, a system with no controls exerted on it. The *stability* of systems in an essential part of control theory. The following statement defines stability of an linear time invariant autonomous system.

Definition 40. A linear system $X(k+1) = \mathbf{A}X(k)$ is said to be stable if

$$\lim_{k \longrightarrow \infty} X(k) = 0$$

starting from an initial state X_0.

The next theorem gives the necessary and sufficient condition required for the stability of autonomous systems.

Theorem 3.1. *An linear time invariant discrete time autonomous system is stable if and only if all the eigen values of* $\mathbf{A}$ *have a magnitude less than 1.*

3.5 Controllability

Consider state transfer from some initial state X_0 to $X(l)$, we say $X(l)$ is reachable in l steps or epochs. The set $\mathbb{R}_l \subseteq \mathbb{R}^n$ defined as:

$$\mathbb{R}_l = \{\sum_{\tau=0}^{l-1} \mathbf{A}^{l-1-\tau}\mathbf{B}U(\tau) : U(\tau) \in \mathbb{R}^m\} \tag{3.12}$$

is the set of reachable state in l epochs.

- $\mathbb{R}_l$ is a subspace of $\mathbb{R}^n$.
- $\mathbb{R}_l \subseteq \mathbb{R}_o$ if $t \leq o$.

Equation 3.9 came rearrange to obtain,

$$X(l) - \mathbf{A}^l X_0 = \mathbf{C}_l \begin{bmatrix} U(l-1) \\ U(l-2) \\ \vdots \\ U(1) \\ U(0) \end{bmatrix} \tag{3.13}$$

where $\mathbf{C}_l = \begin{bmatrix}\mathbf{B}\mathbf{A}\mathbf{B}\cdots\mathbf{A}^{l-1}\mathbf{B}\end{bmatrix}$. Therefore, the reachable set at l, $\mathbb{R}_l = range(\mathbf{C}_l)$. We know from Cayley Hamilton theorem that $\mathbf{A}^k$ for $k \geq n$ can express as a linear combination of $\mathbf{A}^0, \mathbf{A}, \cdots, \mathbf{A}^{n-1}$. Hence for $l \geq n$, $range(\mathbf{C}_l) = range(\mathbf{C}_n)$. Thus we have

$$\mathbb{R}_l = \begin{cases} range(\mathbf{C}_l) & l < n \\ range(\mathbf{C}_n) & l \geq n. \end{cases} \tag{3.14}$$

$\mathbf{C}_n = \mathbf{C}$ is referred in literature as the *controllability* matrix of the system[10]. Note that any state that can be reached can be reached by $l = n$ epochs. In that case the reachable set $\mathbb{R} = range(\mathbf{C})$.

A system is *reachable or controllable* if all the state are reachable (i.e.$\mathbb{R} = \mathbb{R}^n$).
Definition 41. A system is reachable if and only if $Rank(\mathbf{C}) = n$.

Moreover, if a system is controllable then any state can be reached in epochs $\leq n$.

From a robotics perspective, controllability gives insight into whether trajectories are possible. If a system is controllable then we can plan trajectories between any two points in the configuration.

An interesting question to ask at this moment is : how hard is it to steer a system for one state to another? In other words, what is the minimum control effort required to steer a system to some desired state. This leads us to the idea of a *controllability gramian*.

If the system is controllable, particularly, the system of linear equations described in Equation 3.13 has a solution for all $X(l)$, then we would like to compute the minimum norm solution for this system of linear equations. Minimum norm solution for a system of linear equations is a well studied concept in linear algebra for which easy to use formulas are available [26]. The problem can also be framed as the following optimization problem.

$$\min \sum_{\tau=0}^{l-1} \|U(\tau)\|^2 \tag{3.15}$$

subject to:

$$X(l) - \mathbf{A}^l X_0 = \mathbb{C}_l \begin{bmatrix} U(l-1) \\ U(l-2) \\ \vdots \\ U(1) \\ U(0) \end{bmatrix} \tag{3.16}$$

The solution to problem is given by

$$\begin{bmatrix} U_{mn}(l-1) \\ U_{mn}(l-2) \\ \vdots \\ U_{mn}(1) \\ U_{mn}(0) \end{bmatrix} = \mathbb{C}_l^T \left(\mathbb{C}_l \mathbb{C}_l^T\right)^{-1} \bar{X}(l), \tag{3.17}$$

where the subscript mn is used to label the inputs as minimum norm inputs and $\bar{X}(l) = \left(X(l) - \mathbf{A}^l X_0\right)$. The minimum value of the objective Equation 3.15 can be computed in the following manner.

$$\sum_{\tau=0}^{l-1} \|U_{mn}(\tau)\|^2 = \left(\mathbb{C}_l^T \left(\mathbb{C}_l \mathbb{C}_l^T\right)^{-1} \bar{X}(l)\right)^T \left(\mathbb{C}_l^T \left(\mathbb{C}_l \mathbb{C}_l^T\right)^{-1} \bar{X}(l)\right) \tag{3.18}$$

$$= (\bar{X}(l))^T \left(\mathbb{C}_l \mathbb{C}_l^T\right)^{-1} \bar{X}(l) \tag{3.19}$$

$$= (\bar{X}(l))^T \left(\sum_{\tau=0}^{l-1} \mathbf{A}^\tau \mathbf{B}\mathbf{B}^T \left(\mathbf{A}^T\right)^\tau\right)^{-1} \bar{X}(l), \tag{3.20}$$

where $\left(\sum_{\tau=0}^{l-1} \mathbf{A}^\tau \mathbf{B}\mathbf{B}^T \left(\mathbf{A}^T\right)^\tau\right)$ is the finite time *controllability gramian* of the system. The controllability gramian measures the degree of controllability of the system.

3.5.1 Stabilizability

A system is stabilizable if for any $X(0) = X_0$, there exists a $U(k)$ as a feedback law $U(k) = f(X(k))$ such that:

$$\lim_{k \longrightarrow \infty} X(k) = 0. \tag{3.21}$$

Observe that Stabilizability is weaker than controllability.

Theorem 3.2. *If a linear system is controllable then there exist a stabilizing linear feedback* $U(k) = -\mathbf{K}X(k)$.

3.5.2 Pole placement

The basic idea behind the pole placement technique is to construct a linear state feedback gain matrix for a controllable system such that its closed loop eigen values(eigen values of $\mathbf{A} - \mathbf{BK}$) match with some desired values of interest. We will illustrate this idea with an example.

Example 25. *Consider system,*

$$X_{k+1} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} X_k + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \tag{3.22}$$

it can be easily verified that the system is controllable. The open loop poles of the system are -1 *and* -2. *Suppose we wish to design a state feedback gain matrix* $\mathbf{K} = [k_1 k_2]$ *such that closed loop eigenvalues of the system are* 0.5 *and* -0.5 *then observe that the following condition,*

$$det|z\mathbf{I} - (\mathbf{A} - \mathbf{BK})| = z^2 + (3 + k_2)z + (2 + k_1) = 0 \tag{3.23}$$

should be satisfied for all eigen values of the closed loop system. Therefore, by substituting the desired eigen values(0.5, -0.5) in the above equation yields,

$$0.25 + (3 + k_2)0.5 + 2 + k_1 = 0 \tag{3.24}$$

$$0.25 - (3 + k_2)0.5 + 2 + k_1 = 0 \tag{3.25}$$

and solving them we obtain $\mathbf{K} = [-2.25 - 3]$. *Therefore, can be regulated using the control* $u = [2.253]X_k$.

The pole placement technique is easy to implement, but it has a major disadvantage, primarily, how to come up with the desired eigen values for the closed loop system. It would be much easier if the computation of the gain

matrix is based on a more intuitive quantity like energy of the control. This idea is the basis for the development of *Linear quadratic regulator* or LQR.

Although, state feedback is an useful and powerful method for control design, it is usually not possible to measure the full state of a system through its sensors. Therefore, it is required to build a filter or an estimator system which can estimate the full state of the system using its outputs. In literature, such a system is commonly referred as a state estimator or an observer. The common types of observers are luenberger observer and kalman filter. Next, we will examine the conditions required for a system to build a good observer for it.

3.6 Observability

Definition 42. (Observability). A system is said to be observable if, for any initial state $X(0)$ and for any known sequence of inputs $U(0), U(1), \cdots$, there is a positive integer l such that $X(0)$ can be recovered from the outputs $Y(0), Y(1), \cdots, Y(l)$.

If we iterate output equation of the linear time invariant system yields the following matrix equation,

$$\begin{bmatrix} Y(0) \\ Y(1) \\ Y(2) \\ \vdots \\ Y(l) \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^l \end{bmatrix} X(0) + \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{l-1}\mathbf{B} & \mathbf{CA}^{l-2}\mathbf{B} & \mathbf{CA}^{l-3}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix} \begin{bmatrix} U(0) \\ U(1) \\ U(2) \\ \vdots \\ U(l) \end{bmatrix} \tag{3.26}$$

It is straightforward to see from the above matrix equation that observability depends on the matrix $[\mathbf{C}^T; (\mathbf{CA})^T; (\mathbf{CA}^2)^T; \cdots (\mathbf{CA}^l)^T]^T$ which is referred as the *observability matrix* denoted using the symbol $\mathbb{O}_l$. The matrix which is multiplied with the input sequence in Equation 3.26 is commonly referred in literature as *invertibility matrix*, which gives the conditions to recover inputs from outputs of a system given the initial state.

Theorem 3.3. *A necessary and sufficient condition for the system to be observable is that the null space (or kernel) of the observability matrix contains only the zero element.*

Alternately, a system is observable if and only if $rank(\mathbb{O}_{n-rank(\mathbf{C})}) = n$.

We can also analysis other inverse problems associated with the system such as 1) given an output sequence and initial condition can we uniquely identify the inputs which lead to the outputs(Invertibility), 2) given an output sequence and associated inputs can we uniquely identify the initial condition (strong observability). In depth discussion on these inverse problems are beyond on the scope of this course.

Observability gives us the conditions which are essential for building a state estimator or filter. If the system is unobservable then it is impossible to accurately estimate the state of the system with any kind of state estimator or filter (e.g luenberger observer, kalman filter).

Similar to controllability matrix, the observability matrix only answer a "yes" or "no" query, meaning it tell whether the system is observable or not. Akin to controllability, we also define a quantity called the *observability gramian* which measures the ability of sensor system to accurately estimate the initial state.

3.6.1 Observability gramian

Without loss of generality, consider the output sequence $Y(0), Y(1), \cdots Y(l-1)$ obtained from an autonomous system, then the energy of the output sequence defined as $\sum_{\tau=0}^{l-1} \|Y(\tau)\|^2$ can be expressed as,

$$\sum_{\tau=0}^{l-1} \|Y(\tau)\|^2 = \begin{bmatrix} Y(0)^T & Y(1)^T & Y(2)^T \cdots & Y(l)^T \end{bmatrix} \begin{bmatrix} Y(0) \\ Y(1) \\ Y(2) \\ \vdots \\ Y(l) \end{bmatrix} \tag{3.27}$$

$$= \left(\mathbb{O}_l X(0)\right)^T \left(\mathbb{O}_l X(0)\right) \tag{3.28}$$

$$= X(0)^T \left(\mathbb{O}_l^T \mathbb{O}_l\right) X(0) \tag{3.29}$$

$$= X(0)^T \left(\sum_{\tau=0}^{l} \left(\mathbf{A}^T\right)^\tau \mathbf{C}^T \mathbf{C} \mathbf{A}^\tau \right) X(0), \tag{3.30}$$

where $\sum_{\tau=0}^{l} \left(\mathbf{A}^T\right)^\tau \mathbf{C}^T \mathbf{C} \mathbf{A}^\tau$ is the observability gramian of the system. Observe that the observability gramian describe the energy distribution of the output sequence, meaning, it quantifies how hard it is estimate the initial state in a direction compared to another. Particularly, it is much easier to estimate an initial state aligned with the eigen vector of the observability gramian associated with the highest eigen value compared to the one connected with lowest eigen value.

3.7 Duality

For every linear system Equation 3.3 there exist a *dual system* given by,

$$Z(k+1) = \mathbf{A}^T Z(k) + \mathbf{C}^T U(k), \quad X(0) = X_0 \tag{3.31}$$

$$Y(k) = \mathbf{B}^T Z(k) + \mathbf{D} U(k), \quad k = 0, 1, 2 \cdots \tag{3.32}$$

Theorem 3.4. *The system is controllable (observable) if and only if the dual system in observable (controllable).*

4 Nonlinear systems

As mentioned earlier, almost every realistic system is non linear in nature. Therefore, it is important to have a reasonable understanding of non linear system theory to analyze and design control laws for real world systems. As discussed in the previous lecture, the first approach to analyzing a non linear system is through linearization(details of this approach will be discussed in a subsequent section). In fact, control design for most of the aircrafts we travel are carried out using linearization techniques. There are primarily two drawbacks for the linearization method. First, the linearized model are a good approximation for the non linear system only in a local neighborhood around an operating point. Secondly, there are non linear system phenomena which cannot be predicted or described by a linear model. Some such non linear phenomena are:

- **Finite time escape** The state of a non linear system can go to infinity in finite time. However, even the state of an unstable linear system approaches infinity only asymptotically. For example $\dot{x} = -x^2$.
- **Multiple isolated equilibrium points**. In a linear system, there can be one equilibrium point or infinitely many continuous equilibrium points. Whereas, a non linear may contain multiple isolated equilibrium points. Example dynamics of a simple pendulum.
- **Limit cycles**. The amplitude of oscillation in a linear system depends on the initial state of the system. But a non linear system can generate oscillations with fixed amplitude and frequency independent of its initial conditions. Example: Van der Pol oscillator.
- **Chaos**. Non linear system sometimes exhibits strange steady state behaviors which are neither an equilibrium nor a limit cycle. These chaotic behaviors sometimes display randomness, whilst the system is deterministic. Example: Lorenz attractor.

4.1 Existence and uniqueness of solutions

Consider the continuous time non linear system,

$$\dot{X} = F(t, X),\ X(0) = X_0, \tag{4.1}$$

we now examine the conditions for this system to an have unique solution.

Theorem 4.1. ***(Local Existence and Uniqueness:)***. *Let $F(t, X)$ be a piecewise continuous function in t and satisfy the Lipschitz condition*

$$\|F(t, X) - F(t, Y)\| \leq L\|X - Y\|$$

$\forall X, Y \in B(X_0, r), \forall t \in [0, t_1]$. *Then there exist some $\delta > 0$ such that Equation 4.1 has an unique solution over* $[0, \delta]$.

Proof. Refer [13]. □

The theorem can be extended for existence of global solution.

Theorem 4.2. ***(Local Existence and Uniqueness:)***. *Let $F(t, X)$ be a piecewise continuous function in t and satisfy the Lipschitz condition*

$$\|F(t, X) - F(t, Y)\| \leq L\|X - Y\|$$

$\forall X, Y \in \mathbb{R}^n, \forall t \in [0, t_1]$. *Then Equation 4.1 has an unique solution over* $[0, t_1]$.

Proof. Refer [13]. □

Lemma 1. *If $F(t, X)$ and its partial derivatives $\frac{\partial f_i}{\partial x_j}$ are continuous for all X, then $F(t, X)$ is globally Lipschitz in X if and if only the partial derivatives $\frac{\partial f_i}{\partial x_j}$ are globally bounded, uniformly in t.*

The results in the continuous time system can be also extended to discrete time systems $X_{k+1} = F(X_k)$. But in this section we will be discussing primarily about continuous time systems.

4.2 Stability of autonomous systems

Consider the autonomous system

$$\dot{X} = F(X), \tag{4.2}$$

the equilibrium points are the real solutions for the equation

$$F(X) = 0. \tag{4.3}$$

Example 26. *A linear system $\dot{X} = \mathbf{A}X$ can have an isolated equilibrium point at $X = 0$(if $\mathbf{A}$ is non singular) or a continuum of equilibrium points in the null space of $\mathbf{A}$ (if $\mathbf{A}$ is singular).*
Example 27. *Consider the dynamics equation of a simple pendulum,*

$$\dot{x}_1 = x_2 \tag{4.4}$$
$$\dot{x}_2 = -a\sin(x_1) - bx_2,\ \ a, b > 0. \tag{4.5}$$

The equilibrium point of the system computed by solving

$$x_2 = 0 \tag{4.6}$$
$$-a\sin(x_1) - bx_2 = 0. \tag{4.7}$$

The equilibrium points are $(n\pi, 0)$ for $n = 0, \pm 1, \pm 2, \cdots$

In the case of discrete time non linear system evolving according to $X_{k+1} = F(X_k)$, the concept of equilibrium points is replaced using the idea of fixed points. X^* is a fixed point of $X_{k+1} = F(X_k)$ if $X^* = F(X^*)$.

We now enlist the different notions stability used for analyzing non linear systems. Without loss of generality, we assume that $X = 0$ is an equilibrium point of Equation 4.2.
Definition 43. The system is *Lyapunov stable* if for each $\epsilon > 0$ there exists a $\delta > 0$ such that,

$$\|X(0)\| < \delta \implies \|X(t)\| < \epsilon, \quad \forall t > 0. \tag{4.8}$$

Definition 44. The system is *asymptotically stable* if it is Lyapunov stable there exists a $\delta > 0$ such that,

$$\|X(0)\| < \delta \implies \lim_{t \longrightarrow \infty} \|X(t)\| = 0. \tag{4.9}$$

4.2.1 Lyapunov's indirect method

Theorem 4.3. *Let $X = X_e$ be an equilibrium point for Equation 4.2 where F is continuously differentiable around a neighborhood $\mathbb{D}$ about X_e, then*

$$\mathbf{A} = \left.\frac{\partial F}{\partial X}\right|_{X=X_e}$$

then,

- *X_e is asymptotically stable if $Re(\lambda_i) < 0$ for all eigen values of $\mathbf{A}$.*
- *X_e is exponentially stable if and only if $Re(\lambda_i) < 0$ for all eigen values of $\mathbf{A}$.*
- *X_e is unstable if $Re(\lambda_i) > 0$ for any eigen values of $\mathbf{A}$.*

Example 28. *Consider the linearization of Equation 4.4 at $X_e = 0$ with $a = b = 4$,*

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -4 & -4 \end{bmatrix}$$

The eigen values of $\mathbf{A}$ are -2 and -2. Therefore, 0 is a stable equilibrium point.

4.2.2 Lyapunov's direct method

For the remaining part of the article unless otherwise stated, without loss of generality, we assume that 0 is an equilibrium point of the system.

Theorem 4.4. *Let 0 be an equilibrium point of Equation 4.2 and $\mathbb{D} \subset \mathbb{R}^n$ be an open set containing 0. Let $V : \mathbb{D} \longrightarrow \mathbb{R}$ be an continuously differentiable function(Lyapunov function) such that*

$$V(0) = 0 \tag{4.10}$$

$$V(X) > 0 \ \forall X \in \mathbb{D} - 0 \tag{4.11}$$

$$\dot{V}(X) \leq 0 \ \forall X \in \mathbb{D} \tag{4.12}$$

Then, $X = 0$ is Lyapunov stable. In addition if,

$$\dot{V}(X) < 0, \tag{4.13}$$

then $X = 0$ is asymptotically stable.

Proof. Refer [13]. □

In order to extend the theorem to understand global asymptotic stability of the origin we need the Lyapunov function to be radially unbounded.

Example 29. *Consider the system,*

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}-x_1 + 2x_1^2x_2\\ -x_2\end{bmatrix} \tag{4.14}$$

and the Lyapunov function,

$$V(X) = x_1^2 + x_2^2. \tag{4.15}$$

Then,

$$\dot{V}(X) = \nabla V \cdot F(x) \tag{4.16}$$

$$= -2x_2^2 - 2x_1^2(1 - 2x_1x_2). \tag{4.17}$$

Note that, $\dot{V}(X) < 0$ when $(1 - 2x_1x_2) > 0$ or $x_1x_2 < \frac{1}{2}$. Therefore, the origin is locally asymptotically stable.

Definition 45. Domain of attraction [25].The domain of attraction of an equilibrium point X_0 is the set,

$$\mathbb{D}(X_0) = \{X \in \mathbb{R}^n : \lim_{t \longrightarrow \infty} \phi(t, X) = X_0\}, \tag{4.18}$$

where $\phi(t, X)$ is the state of the system at time t after starting from X at time zero.

4.3 State feedback stabilization

Consider the general non linear system $\dot{X} = F(t, X, U)$, the problem of state feedback stabilization can be defined as the constructing a state feedback law

$$U = \Gamma(t, X) \tag{4.19}$$

such that the origin is an uniformly asymptotically stable equilibrium point for the closed loop system $\dot{X} = F(t, X, \Gamma(t, X))$. Although, the original definition is concerned with the stabilization of origin, the definition can be extended for other points of interest.

4.3.1 Linear feedback stabilization

Suppose, we want stabilize $\dot{X} = F(X, U)$ at a point $X = X_d$ and we were able to compute a $U = U_d$ such that X_d is an equilibrium point or

$$F(X_d, U_d) = 0, \tag{4.20}$$

then we can stabilize the system locally if the linearized of system about (X_d, U_d) is controllable. Let $X = X_d + X_\delta$ and $U = U_d + U_\delta$ then,

$$\dot{X} = \dot{X}_d + \dot{X}_\delta = F(X_d + X_\delta, U = U_d + U_\delta).$$

Using taylor series expansion we obtain,

$$\dot{X}_\delta = F(X_d, U_d) + \left.\frac{\partial F}{\partial X}\right|_{X_d,U_d} X_\delta + \left.\frac{\partial F}{\partial U}\right|_{X_d,U_d} U_\delta \tag{4.21}$$

$$\dot{X}_\delta = \left.\frac{\partial F}{\partial X}\right|_{X_d,U_d} X_\delta + \left.\frac{\partial F}{\partial U}\right|_{X_d,U_d} U_\delta \tag{4.22}$$

Now we can use the tools from linear system theory to stabilize the non linear system. If we compute a state feedback gain matrix $\mathbf{K}$ using method like pole placement or LQR, then we can use feedback law $U = U_d + \mathbf{K}(X - X_d)$ to stabilize the non linear system. The idea can be easily applied to trajectory stabilization(tracking) of a nominal trajectory $(X_0(t), U_0(t))$ such that $\dot{X}_0(t) = F(X_0(t), U_0(t))$. This is a widely used technique for control design.

4.3.2 Gain scheduling

In the previous subsection, we looked at a design method that guarantee stability in some neighborhood of the equilibrium point. However, this is very limiting as we would like to stabilize the system over a large region of the state space about several equilibrium points. This can be achieved by linearizing the system about each equilibrium point and design a local controller for stability. Then, we can online schedule the gains to go from one equilibrium point to another. This design methodology which scales the previous design methods can be described by the following steps:

1. Linearize the given nonlinear system about several operating points, which are parametrized by scheduling variables.
2. Design a parametrized family of linear controllers to locally stabilize the system around each of the operating points.
3. Develop a gain-scheduled controller.
4. Test the performance of the gain-scheduled controller through simulation of the nonlinear closed-loop model.

Consider the non linear system and an associated output equation,

$$\dot{X} = F(X, U) \tag{4.23}$$

$$y = H(X). \tag{4.24}$$

Suppose we would to this system to track a reference signal r. It is clear that unlike in a linear system, a linear controller with a common set of gains cannot be used to track any reference signal of choice. The gain may have to vary based on the reference signal used. Now assume that there exist an unique pair (X_d, U_d) such that,

$$0 = F(X_d(r), U_d)(r) \tag{4.25}$$

$$r = H(X_d)(r) \tag{4.26}$$

for all choices of r. To achieve tracking in the presence of unknown disturbances the controller should have an

integrator. Thus,

$$\dot{e} = y - r \tag{4.27}$$
$$U = \mathbf{K}_1(r)X - k_2 e. \tag{4.28}$$

The gains $\mathbf{K}_1(r)$ and k_2 are design such that closed loop state matrix

$$\begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{K}_1 & -\mathbf{B}k_2 \\ \mathbf{C} & 0 \end{bmatrix} \tag{4.29}$$

is stable.

Example 30. *Consider the system,*

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \tan(x_1) + x_2 \\ x_1 + u \end{bmatrix} \tag{4.30}$$
$$y = x_2, \tag{4.31}$$

we want track a reference signal r. It is straightforward to observe that,

$$X_d(r) = \begin{bmatrix} -\tan^{-1}(r) \\ r \end{bmatrix}, U_d(r) = \tan^{-1}(r). \tag{4.32}$$

We design the following controller,

$$\dot{e} = y - r \tag{4.33}$$
$$U = \mathbf{K}_1(r)X - k_2 e. \tag{4.34}$$

The linearized matrices of the system are,

$$\mathbf{A} = \begin{bmatrix} 1 + r^2 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{4.35}$$

Finally, we design the gain such that the closed loop system is stable. The computed gains are,

$$\mathbf{K}_1 = [(1 + r^2)(3 + r^2) + 3 + \frac{1}{(1 + r^2)}, 3 + r^2] \tag{4.36}$$
$$k_2 = -\frac{1}{1 + r^2} \tag{4.37}$$

4.3.3 Feedback linearization

In general, stabilization of non linear system is a difficult problem and is still an active area of research. A slightly different type of non linear system called control affine system has attracted a great deal of interest in both the robotics and controls community for eons. This is mostly due to the fact that dynamics of mechanical systems can be written in a control affine way. A non linear system is control affine if it can be written as

$$\dot{X} = F(X) + G(X)U. \tag{4.38}$$

For example, consider the dynamics of a simple pendulum with torque control

$$\dot{x}_1 = x_2 \tag{4.39}$$
$$\dot{x}_2 = -a\sin(x_1) - bx_2 + u,\ \ a, b > 0, \tag{4.40}$$

observe that the dynamics of the system in affine in control variable.

Suppose here is a change of variables $Z = \mathbf{T}(X)$, defined for all $X \in \mathbb{D} \in \mathbb{R}^n$, that transforms the system into the form

$$\dot{Z} = \mathbf{A}Z + \mathbf{B}\gamma(X)[U - \alpha(X)], \tag{4.41}$$

where $(\mathbf{A}, \mathbf{B})$ is controllable and $\gamma(X)$ is non singular for all $X \in \mathbb{D}$. If we assign $U = \alpha(X) + \gamma(X)^(-1)V$ then the above equation reduces to

$$\dot{Z} = \mathbf{A}Z + \mathbf{B}V. \tag{4.42}$$

Therefore, through this feedback process we have converted a non linear system to a linear system. Additionally, since this is a linear system we can use all the tools from linear system theory to design controllers. As an example, the pendulum dynamics Equation 4.39 can be linearized by choosing $u = a\sin(x_1) + v$.

Although, feedback linearization is a useful technique, it suffers from the drawback that it extensively dependent on the dynamics model. In other words, model errors could deeply affect the performance of the controller. Additional, feedback linearization based controller in most cases try to cancel out the natural dynamics of the system in practice it could be highly energy inefficient.

I would like to point out that there is an analogous tool called *differentially flatness*, which is widely used in robotics and control theory. We will not perform a rigorous treatment of the topic here, since it would require to invoke tools from differential geometry. Readers interested in a more formal analysis of the topic may refer to [15]. Crudely speaking, a system is differentially flat if there exist a set of outputs, known as *flat outputs*, for which all states and inputs are determined by the outputs and a finite number of their derivatives [6]. Thus if we specify the trajectory of a system in terms of its flat outputs then we can compute what all of the states/inputs must have been doing. This techniques is extensively used in trajectory planning of mechanical systems like chained carts, quadrotors etc. Note that every feedback linearizable system is differentiable flat. But according to [19, Theorem 2] If a control system is differentially flat then it is dynamic feedback linearizable on an open dense set with the dynamic feedback possibly depending explicitly on time.

5 Optimization/ optimal control

Optimization is the core approach to solving almost every problem that we encounter in engineering. It is a well studied topic in literature and still an active area of research, specifically in engineering, mathematics and physics. But its freshness has never diminished over the years. In general, an optimization problem has the following structure. It may contain a set of constraints over a set represented using a collection of equalities and inequalities. The elements in the set which satisfy these constraints are referred as *feasible solutions* to the problem. In addition, the optimization problem may also admit a cost(reward) function which is to be minimized(maximized) to obtain desired optimal solutions. The optimization framework is also used extensively in control system applications to compute control laws under design constraints. The general terminology given to this framework in control theory literature is *optimal control*. In this section, we will introduce a common technique used in optimization problems known as *Lagrange multipliers*. Subsequently, we will explore the two most important ideas in optimal control namely: *Pontryagin maximum/minimum principle* and *Dynamic programming*. In depth treatment of the concepts discussed in this section can be found in [3, 17]. The reader are directed to [16] to learn more about the computational aspects of optimal control.

5.1 Lagrange multipliers

Consider the problem,

$$\min_{X,U} \; L(X,U) \tag{5.1}$$

subject to,

$$F(X,U) = 0. \tag{5.2}$$

Now, if we define the Lagrangian of the above problem as,

$$\mathbb{L}(X,U,\lambda) = L(X,U) + \lambda F(X,U) \tag{5.3}$$

then the necessary conditions for (X^*, U^*) to be an optimal solution of the problem are,

$$\nabla_X L(X^*,U^*) = \lambda \nabla_X F(X^*,U^*) \tag{5.4}$$
$$\nabla_U L(X^*,U^*) = \lambda \nabla_U F(X^*,U^*) \tag{5.5}$$
$$F(X^*,U^*) = 0. \tag{5.6}$$

Therefore, we could solve Equation 5.4 to find the candidates for optimal solutions. The method of Lagrangian multipliers is very powerful in solving constrained optimization problems.

Example 31. *Find the rectangle of maximum area with perimeter p. That is,*

$$L(x, y) = xy$$

subject to

$$f(x, y) = 2x + 2y - p = 0$$

then using the conditions in Equation 5.4 yields,

$$\begin{aligned} y &= \lambda 2 \\ x &= \lambda 2 \\ 2x + 2y &= p. \end{aligned}$$

By solving the above set of equations we arrive at the optimal solution $x = \frac{p}{4}$ *and* $y = \frac{p}{4}$.

5.2 Optimal control formulation

In this article, we formulate an optimal control problem which tries to address the following problem.

Given a dynamical system $\dot{X} = F(X, U)$, where $X(t) \in \mathbb{R}^n$ is the state of the system and $U : [0, T] \longrightarrow \mathbb{R}^m, T > 0$ is a control action on the system which belongs to a compact set of control actions denoted as $\mathbb{U}$, find a control input which drives the dynamical system from an initial state to some state in time $T > 0$ along with minimizing the cost function

$$J(X, U) = \int_0^T L(X(t), U(t), t)dt + g(X(T)). \tag{5.7}$$

We define this problem formally as,
Problem 1.

$$J(X, U) = \int_0^T L(X(t), U(t), t)dt + g(X(T)) \tag{5.8}$$

Subject to,

$$\dot{X} = F(X, U) \tag{5.9}$$

$$U \in \mathbb{U} \tag{5.10}$$

In Equation 5.7, $\int_0^T L(X(t), U(t))dt$ and $g(X(T), U(T))$ are referred as *running cost* and *terminal cost* respectively.

We can also formulate a discrete time optimal control problem akin to the continuous case. In the discrete case, a discrete dynamical system $X(k + 1) = F(X(k), U(k))$ such that $X(k) \in \mathbb{R}^n$ and $U(k) \in \mathbb{U} \subseteq \mathbb{R}^m$ is consider. Once again, the goal is to compute a sequence of control inputs which can drive the discrete dynamical from one state to some state in a finite number of epochs N.

$$J(X,U) = \sum_{k=0}^{N-1} L(X(k),U(k),k) + g(X(N),N). \tag{5.11}$$

Formally,

Problem 2.

$$J(X,U) = \sum_{k=0}^{N-1} L(X(k),U(k),k) + g(X(N),N). \tag{5.12}$$

Subject to,

$$X(k+1) = F(X(k),U(k)) \tag{5.13}$$

$$U \in \mathbb{U} \tag{5.14}$$

There are other variations of the problem studied in literature namely 1) variable endpoint control problem and 2) fixed time variable endpoint problem. But in this manuscript, we will focus on the above mentioned formulation of the problem.

We will motive this section with a fairly straightforward example.

Example 32. *Consider a particle which follows a double integrator dynamics or in plain language Newton's second law of motion then,*

$$\dot{x} = v \tag{5.15}$$

$$\dot{v} = u, \tag{5.16}$$

such that $u \in [-1,1]$. Our goal is to reach the origin with zero speed(reach the state $x = 0, v = 0$) in minimum time. It is easy to see that the cost function for this problem is,

$$\int_0^T 1dt. \tag{5.17}$$

If we think about it, we can intuitively understand that the solution to problem amounts to using extreme values in the controls set and switching between them at the right moment. This type of control strategy is called the bang bang control. This is a common strategy when using controls with saturation. Lets compute the trajectories of the system at the extreme controls,

1. *If $u = 1$, given $x(0)$ and $v(0)$.*
   Solving the equation of motion we arrive,

$$x(t) = \frac{t^2}{2} + v(0)t + x(0) \tag{5.18}$$

$$v(t) = t + v(0). \tag{5.19}$$

   Furthermore, the phase space((x,v) space) trajectory of the system is $x(t) = \frac{v^2(t)}{2} - c_1(x(0),v(0))$.

2. *If $u = -1$, given $x(0)$ and $v(0)$.*

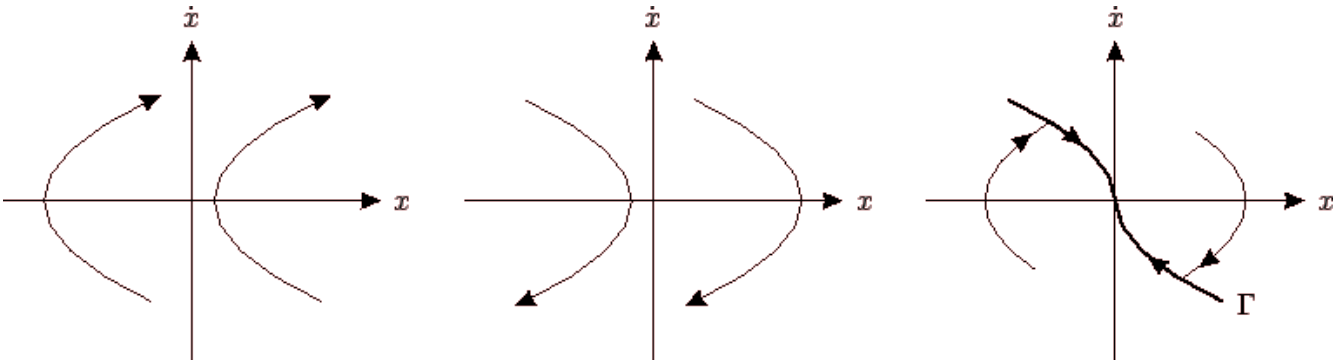


Figure 5.1: Bang-bang time-optimal control of the double integrator: (a) (left)trajectories for $u = 1$, (b)(center) trajectories for $u = -1$, (c)(right) the switching curve and optimal trajectories. Image courtesy: Daniel Liberzon's website.

Similar to the previous case we obtain the equations of motion

$$x(t) = -\frac{t^2}{2} + v(0)t + x(0) \tag{5.20}$$

$$v(t) = -t + v(0) \tag{5.21}$$

and the phase space trajectory $x(t) = -\frac{v^2(t)}{2} - c_2(x(0), v(0))$

These curves are shown in Equation 32(a,b), with the arrows indicating the direction in which they are traversed. It is easy to see that only two of these trajectories hit the origin. Their union is the thick curve in Equation 32(c), which we call the switching curve and denote by Γ*; it is defined by the relation* $x = -\frac{1}{2}|v|v$*. The optimal control strategy thus consists in applying* $u = 1$ *or* $u = -1$ *depending on whether the initial point is below or above* Γ *, then switching the control value exactly on* Γ *and subsequently following* Γ *to the origin; no switching is needed if the initial point is already on* Γ*. We can prove that bang bang control is indeed the optimal control for the double integrator using the tools discussed in this article.*

In the upcoming subsections, we will examine the two main techniques used of solving the optimal control problem.

5.3 Pontryagin minimum principle

Lev Pontryagin was a Russian mathematician, who and his students formulated the problem during the cold war era. Of course, Americans weren't sitting idle, they had their formulation of the problem called *dynamic programming* developed by Richard Bellman. Pontryagin minimum principle is the necessary set of conditions required for global optimality. We will derive the minimum principle for the discrete case. Since derivation of principle for the continuous case relies on calculus of variation which will not discussed in this article, we will only state the conditions for the continuous time case.

We start the derivation using the Lagrangian multiplier technique discussed in subsection 5.1. The Lagrangian of Problem 2 is

$$\mathcal{L}(X_{0:N}, U_{0:N-1}, \lambda_{1:N}) = g(X(N), N) + \sum_{k=0}^{N-1} \left(L(X(k), U(k), k) + \left(F(X(k), U(k)) - X(k+1) \right)^T \lambda_{k+1} \right), \tag{5.22}$$

where $\lambda_{0:N} = (\lambda_0, \lambda_1, \cdots, \lambda_N)$ denotes the vector of Lagrange multipliers. We define the Hamiltonian for Problem 2 to be

$$\mathcal{H}^k(X, U, \lambda) = L(X, U, k) + F(X, U)^T \lambda \tag{5.23}$$

and rearrange the terms in Equation 5.22 to obtain,

$$\mathcal{L}(X_{0:N}, U_{0:N-1}, \lambda_{1:N}) = g(X(N), N) - X(N)^T \lambda_N + X(0)^T \lambda_0 + \sum_{k=0}^{N-1} \left(\mathcal{H}^k(X(k), U(k), \lambda_{k+1}) - X(k)^T \lambda_k \right) \tag{5.24}$$

Now examine the differential of $\mathcal{L}$ and contemplate its changes with respect to the changes in U and X. We have

$$d\mathcal{L} = \left(\frac{\partial g(X(N), N)}{\partial X(N)} - \lambda_N \right)^T dX(N) + \lambda_0^T dX(0) + \sum_{k=0}^{N-1} \left(\left(\frac{\partial \mathcal{H}^k}{\partial X(k)} - \lambda_k \right)^T dX(k) + \left(\frac{\partial \mathcal{H}^k}{\partial U(k)} \right)^T dU(k) \right).$$

To make the variations along $dX(k)$ zero we need $\frac{\partial \mathcal{L}}{\partial X(k)} = 0$ and choose the following Lagrange multipliers to satisfy this condition.

$$\lambda_k = \frac{\partial \mathcal{H}^k}{\partial X(k)} = \frac{\partial}{\partial X(k)} L(X(k), U(k), k) + \left(\frac{\partial F(X(k), U(k))}{\partial X(k)} \right)^T \lambda_{k+1}, \ 0 \leq k < N \tag{5.25}$$

$$\lambda_N = \frac{\partial g(X(N), N)}{\partial X(N)} \tag{5.26}$$

By making this choice $d\mathcal{L}$ can be expressed as

$$d\mathcal{L} = \lambda_0 dX(0) + \sum_{k=0}^{N-1} \left(\left(\frac{\partial \mathcal{H}^k}{\partial U(k)} \right)^T dU(k) \right). \tag{5.27}$$

If the initial condition fixed, then $dX(0) = 0$. The second term in Equation 5.27 is zero $U(k)$ is the unconstrained minimum of $\mathcal{H}^k$ ($\frac{\partial \mathcal{H}^k}{\partial U(k)} = 0$). Final, summarizing the conditions yields the minimum principle,

$$X(k+1) = F(X(k), U(k)) \tag{5.28}$$

$$\lambda_k = \frac{\partial}{\partial X(k)} L(X(k), U(k), k) + \left(\frac{\partial F(X(k), U(k))}{\partial X(k)} \right)^T \lambda_{k+1}, \ 0 \leq k < N \tag{5.29}$$

$$\lambda_N = \frac{\partial g(X(N), N)}{\partial X(N)} \tag{5.30}$$

$$U(k) = \underset{U \in \mathbb{U}}{argmin} \mathcal{H}^k(X(k), U, \lambda_k + 1), \tag{5.31}$$

for a given $X(0)$ and N.

Given an initial state $X(0)$ and final time t_f, the minimum principle in the continuous time scenario can be stated as

$$\dot{X}(t) = \frac{\partial}{\partial P}\mathcal{H}^k(X(t), U(t), P(t), t) \tag{5.32}$$

$$\dot{P}(t) = -\frac{\partial}{\partial X}\mathcal{H}^k(X(t), U(t), P(t), t) \tag{5.33}$$

$$P(t_f) = \frac{\partial g(X(t_f), t_f)}{\partial X(t_f)} \tag{5.34}$$

$$U(t) = \underset{U \in \mathbb{U}}{argmin}\mathcal{H}^k(X(t), U, P(t), t), \tag{5.35}$$

where $\mathcal{H}^k$ is the Hamiltonian function defined as

$$\mathcal{H}^t(X(t), U(t), P(t), t) = L(X(t), U(t), t) + F(X(t), U(t))^T P(t) \tag{5.36}$$

and $P(t)$ is termed the *costate*. The second equation in Equation 5.32 and Equation 5.28 is referred to as the *adjoint equation*. It is also called *back propagation* in some other communities.

Using the minimum principle we can prove that the bang bang control is the optimal policy for Example 32. The readers may refer to [17] for details. In addition, for linear systems minimum principle becomes the sufficient condition when the control set is convex.

5.3.1 Numerical implementation

From Equation 5.27, it is straightforward to understand that

$$\frac{\partial \mathcal{H}^k}{\partial U} = \left[\frac{\partial \mathcal{H}^0}{\partial U(0)}, \frac{\partial \mathcal{H}^1}{\partial U(1)}, \cdots, \frac{\partial \mathcal{H}^{N-1}}{\partial U(N-1)}\right] \tag{5.37}$$

$$\tag{5.38}$$

is the gradient of the total cost function with respect to the control input, where $\frac{\partial \mathcal{H}^k}{\partial U(k)} = \frac{\partial}{\partial U(k)}L(X(k), U(k), k) + \frac{\partial}{\partial U(k)}F(X(k), U(k))^T\lambda_{k+1}$.

Since we have the gradient of the cost function with respect to control inputs we can use a gradient decent strategy to compute the optimal control. The steps for the gradient decent algorithm are as follows,

1. Given a sequence of inputs, compute the sequence states by evaluating the dynamics equation forward in time. Then, solve the adjoint equation in Equation 5.28 backward time to compute the costates ($\lambda_{0:N}$).
2. Compute the gradient using Equation 5.37 and improve the control inputs using gradient decent. Go to step 1 until convergence.

5.4 Dynamic programming

Dynamic programming is a consequence of a particular mathematical wisdom: "It is sometimes easier to solve a problem by embedding it within a larger class of problems and then solving the larger class all at once." To motivate this consider this example.

Example 33. *Consider the following integral,*

$$\int_0^\infty \frac{\sin x}{x} dx. \tag{5.39}$$

This not an easy integral to compute directly, so let's consider a different problem,

$$I(a) = \int_0^\infty e^{-ax} \frac{\sin x}{x} dx. \tag{5.40}$$

The solution for the above integral is $I(a) = -\tan^{-1}(a) + \frac{\pi}{2}$. *Thus we can conclude that,*

$$\int_0^\infty \frac{\sin x}{x} dx = I(0) = \frac{\pi}{2}. \tag{5.41}$$

A similar idea when adapted for optimal control problems yield the framework of *dynamic programming*. In the context of optimal control instead of solving the problem for an initial state, we solve a more general problem. In the more general contest, we consider solving the problem by varying initial states and initial times, for which our problem is a special case. We will start by deriving the equations associates with discrete time problem (Problem 2). The dynamic programming is also used in the context of graph searching. In fact, if we consider the discrete space problem of (Problem 2) it is precisely a graph searching problem. In practice, many discrete time optimal control problems are solved by discretizing the state space and preforming a graph search.

We start the derivation by defining a quantity called the *value function* or the *cost to go* $\mathcal{V}(X,t)$. $\mathcal{V}(X,k)$ outputs the optimal cost to go starting from state X at time step k according to the cost function Equation 5.11 while complying to discrete dynamics $X(k+1) = F(X(k), U(k))$. Thus, given an

$$V(X,0) = \min_{U_{0:N-1} \in \mathbb{U}} \sum_{k=0}^{N-1} L(X(k), U(k), k) + g(X(N), N), \;\; X(0) = X \tag{5.42}$$

$$V(X,N) = g(X(N), N), \;\; X(N) = X \tag{5.43}$$

$$V(X,N-1) = \min_{U_{N-1} \in \mathbb{U}} L(X(N-1), U(N-1), N-1) + g(X(N), N), \;\; X(N-1) = X \tag{5.44}$$

Due to additive structure of the cost function, we observe that the cost to go function can be written in a recursive fashion. Hence,

$$V(X,k) = \min_{U \in \mathbb{U}} \{L(X,U,k) + V(F(X,U), k+1)\} \tag{5.45}$$

$$V(X,N) = g(X,N). \tag{5.46}$$

Equation 5.45 is in accordance with the *principle of optimality*. Principle of Optimality states that: An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision [2]. In other words, If we have an optimal trajectory starting from an initial state $X(0)$ and if we remove the part of the trajectory starting from $X(0)$ until some state $X(t_1)$, then rest of the trajectory starting $X(t_1)$ is also optimal.

Now using cost to go formulation we can compute an optimal control for the problem as

$$U^*(X,k) = \underset{U\in\mathbb{U}}{argmin}\{L(X,U,k) + V(F(X,U),k+1)\} \tag{5.47}$$

where $U^*(X,k)$ is the optimal feedback control or policy.

Again by defining a value function and using the principle of optimality one could derive a formula for the optimal policy for the continuous case (Problem 1). Next we will define the appropriate value function and state the conditions for optimality. Readers may refer to [3, 17] for the complete derivation. The value function for Problem 1 can be defined as

$$\mathcal{V}(t,X) = \underset{U_{[t,T]}\in\mathbb{U}}{min}\left\{\int_t^T L(X(\tau),U(\tau),\tau)d\tau + g(X(T))\right\},\ X(t) = X \tag{5.48}$$

$$\mathcal{V}(T,X) = g(X) \tag{5.49}$$

where the notation $U_{[t,T]}$ indicates that the control U is restricted to the interval $[t,T]$. It can be shown that the value function should satisfy the following partial differential equation $\forall t \in [0,T]$

$$\frac{\partial \mathcal{V}}{\partial t} + \underset{U\in\mathbb{U}}{min}\left\{L(X,U,t) + \left\langle \frac{\partial \mathcal{V}}{\partial X}, F(X,U)\right\rangle\right\} = 0. \tag{5.50}$$

The above equation for the value function is called the **Hamilton-Jacobi-Bellman (HJB)** equation. The associated boundary condition with HJB equation is $\mathcal{V}(T,X) = g(X)$. Also, the optimal feedback control policy can be computed using

$$U^* = \underset{U\in\mathbb{U}}{argmin}\left\{L(X,U,t) + \left\langle \frac{\partial \mathcal{V}}{\partial X}, F(X,U)\right\rangle\right\} \tag{5.51}$$

The next theorem states that HJB is a sufficiency condition.

Theorem 5.1. *(HJB) sufficiency theorem. Suppose that a $\mathcal{C}^1$ function $\widehat{\mathcal{V}} : [t_0,t_1]\times\mathbb{R}^n \longrightarrow \mathbb{R}$ satisfies the HJB equation*

$$\frac{\partial \widehat{\mathcal{V}}}{\partial t} + \underset{U\in\mathbb{U}}{min}\left\{L(X,U,t) + \left\langle \frac{\partial \widehat{\mathcal{V}}}{\partial X}, F(X,U)\right\rangle\right\} = 0. \tag{5.52}$$

(for all $t \in [0,T)$ and all $X \in \mathbb{R}^n$) and the boundary condition $\widehat{\mathcal{V}}(T,X) = g(X)$.

Suppose that a control $\hat{U} : [t_0, t_1] \longrightarrow \mathbb{U}$ *and the corresponding trajectory* $\hat{X} : [0, T] \longrightarrow \mathbb{R}^n$ *, with the given initial condition* $\hat{x}(0) = X_0$ *, satisfy everywhere the equation*

$$\left\{ L(\hat{X}, \hat{U}, t) + \left\langle \frac{\partial \widehat{\mathcal{V}}(t, \hat{X})}{\partial X}, F(\hat{X}, \hat{U}) \right\rangle \right\} = \min_{U \in \mathbb{U}} \left\{ L(\hat{X}, U, t) + \left\langle \frac{\partial \widehat{\mathcal{V}}(t, \hat{X})}{\partial X}, F(\hat{X}, U) \right\rangle \right\} \tag{5.53}$$

Then $\widehat{V}(0, X_0)$ *is the optimal cost (i.e.,* $\widehat{V}(0, X_0) = V(0, X_0)$ *where* V *is the value function) and* $\hat{U}$ *is an optimal control.*

In other words, any policy satisfying HJB equation is an optimal policy. Note that this optimal control is not claimed to be unique; there can be multiple controls giving the same cost. Therefore, HJB equation is necessary and sufficient condition for optimality when solved over the whole state space for an optimum. Also observe HJB is a nice tool to verify if your policy is optimal. We illustrate this with an example [28].

Example 34. *Consider a double integrator dynamics problem with following cost function:*

$$\int_0^\infty L(x(t), v(t), u(t))dt = \int_0^\infty (x(t))^2 + (v(t))^2 + (u(t))^2 dt. \tag{5.54}$$

Lets assume that someone made a claim that

$$u(x, v) = -x - \sqrt{3}v \tag{5.55}$$

is an optimal control for this objective function. Now to convince you that this indeed an optimal control, the person has to use the value function,

$$V(x, v) = \sqrt{3}x^2 + 2xv + \sqrt{3}v^2 \tag{5.56}$$

and compute,

$$\frac{\partial V}{\partial x} = 2\sqrt{3}x + 2v, \quad \frac{\partial V}{\partial v} = 2\sqrt{3}v + 2x.$$

Therefore,

$$L(x(t), v(t), u(t)) + \frac{\partial V}{\partial x}v + \frac{\partial V}{\partial v}u = x^2 + v^2 + u^2 + (2\sqrt{3}x + 2v)v + (2\sqrt{3}v + 2x)x. \tag{5.57}$$

The minimum of above function about u *can be computed by taking derivatives about* u *and setting it to zero, which yields,*

$$u = -x - \sqrt{3}v.$$

Therefore, proposed controller is optimal.

Now, we show that LQR is optimal by deriving it you dynamic programming[16].

Example 35. *(LQR). Consider the discrete time dynamics* $X(k+1) = \mathbf{A}X(k) + \mathbf{B}U(k)$ *and cost function*

$$J(X, U) = \frac{1}{2} \sum_{k=0}^{N-1} X(k)^T \mathbf{Q} X(k) + U(k)^T \mathbf{R} U(k) + \frac{1}{2} X(N)^T \mathbf{S}(N) X(N), \tag{5.58}$$

where $\mathbf{Q} \geq 0$, $\mathbf{R} > 0$ *and* $\mathbf{S}(N) \geq 0$. *If* $V(X,k)$ *is the cost-to go function then,*

$$V(X,N) = \frac{1}{2} X^T \mathbf{S}(N) X^T \tag{5.59}$$

$$V(X,N-1) = X^T \mathbf{Q} X + U(N-1)^T \mathbf{R} U(N-1) + \frac{1}{2} \left(\mathbf{A}X + \mathbf{B}U(N-1)\right)^T \mathbf{S}(N) \left(\mathbf{A}X + \mathbf{B}U(N-1)\right). \tag{5.60}$$

Since the controls are unconstrained, the control can be compute by differentiation,

$$\frac{\partial V}{\partial U(N-1)} = 0 = \mathbf{R}U(N-1) + \mathbf{B}^T \mathbf{S}(N) \left(\mathbf{A}X + \mathbf{B}U(N-1)\right). \tag{5.61}$$

Thus,

$$U(N-1) = -\left(\mathbf{B}^T \mathbf{S}(N) \mathbf{B} + \mathbf{R}\right)^{-1} \mathbf{B}^T \mathbf{S}(N) \mathbf{A} X(N-1),\ X(N-1) = X. \tag{5.62}$$

If we define,

$$\mathbf{K}(N-1) = \left(\mathbf{B}^T \mathbf{S}(N) \mathbf{B} + \mathbf{R}\right)^{-1} \mathbf{B}^T \mathbf{S}(N) \mathbf{A}, \tag{5.63}$$

then our optimal policy is,

$$U^*(N-1) = -\mathbf{K}(N-1) X(N-1). \tag{5.64}$$

By substituting Equation 5.64 in cost to go, we derive the optimal cost to go function as,

$$V^*(X,N-1) = \frac{1}{2} X^T \left[\mathbf{K}^T(N-1) \mathbf{R} \mathbf{K}(N-1) + \mathbf{Q} + \left(\mathbf{A} - \mathbf{B}\mathbf{K}(N-1)\right)^T \mathbf{S}(N) \left(\mathbf{A} - \mathbf{B}\mathbf{K}(N-1)\right)\right] X. \tag{5.65}$$

Also, by defining,

$$\mathbf{S}(N-1) \triangleq \mathbf{K}^T(N-1) \mathbf{R} \mathbf{K}(N-1) + \mathbf{Q} + \left(\mathbf{A} - \mathbf{B}\mathbf{K}(N-1)\right)^T \mathbf{S}(N) \left(\mathbf{A} - \mathbf{B}\mathbf{K}(N-1)\right), \tag{5.66}$$

we can express the optimal cost to go as,

$$V^*(X,N-1) = \frac{1}{2} X(N-1)^T \mathbf{S}(N-1) X(N-1). \tag{5.67}$$

We can observe this pattern for every time step backwards in time$(N-2, N-3, \cdots, 0)$*. Therefore, the whole optimal solution can be succinctly written as,*

$$\mathbf{K}(k) = \left(\mathbf{B}^T\mathbf{S}(k+1)\mathbf{B} + \mathbf{R}\right)^{-1}\mathbf{B}^T\mathbf{S}(k+1)\mathbf{A} \tag{5.68}$$

$$U^*(k) = -\mathbf{K}(k)X(k) \tag{5.69}$$

$$\mathbf{S}(k) = \mathbf{K}^T(k)\mathbf{R}\mathbf{K}(k) + \mathbf{Q} + (\mathbf{A} - \mathbf{B}\mathbf{K}(k))^T\mathbf{S}(k+1)(\mathbf{A} - \mathbf{B}\mathbf{K}(k)) \tag{5.70}$$

$$V^*(X(k), k) = \frac{1}{2}X(k)^T\mathbf{S}(k)X(k) \tag{5.71}$$

where Equation 5.70 is solved backward from the given $S(N)$ and is referred in literature as discrete time Riccati equation. One can find a striking resemblance of these equation with the Kalman filter equations. In fact, Kalman filter is can also be derived in the same way.

Figure 6.1: Homeomorphism between a doughnut and a coffee cup

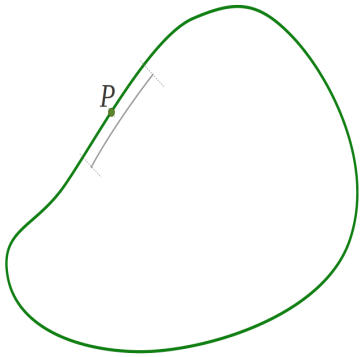


Figure 6.2: A circle is a 1-manifold. A open neighborhood of every point on it resembles an open segment of $\mathbb{R}$.

6 Differential Geometry

[1]How can we compute quantities like velocity vectors, gradient without leaving the space your are in? In real life we mostly compute things while staying on earth without realizing that our earth is embedded in a 3D space(Physicists might disagree with me on this). Differential geometry is all about formalizing this idea of computing quantities in an intrinsic manner.

6.1 Manifold

Definition 46. (Homeomorphism. [20] Figure 6.1). Two topological spaces $\mathbb{X}$ and $\mathbb{Y}$ are *homeomorphic* if there exists a bijective function $f : \mathbb{X} \longrightarrow \mathbb{Y}$ (which implies the inverse, $f^{-1} : \mathbb{Y} \longrightarrow \mathbb{X}$ exists and is bijective) such that both f and f^{-1} and continuous.

Definition 47. (Manifold [5] Figure 6.2). A manifold is a topological space that locally looks like an Euclidean space everywhere. That is, if $\mathbb{M}$ is a topological space, and $p \in \mathbb{M}$ is a point in it, then there exists a open neighborhood $\mathbb{U}$ of p (i.e. an open set $\mathbb{U}$ containing p)), such that one can construct homeomorphisms $\psi : \mathbb{U} \longrightarrow \mathbb{R}^d$ for some nonnegative integer d. The minimum value of D for which it is possible to construct such homeomorphisms is called dimension of the manifold.

Of course all manifolds are topological spaces, but the converse is not true. A circle is a one dimensional manifold

[1]Thanks to my friend Subhrajit Bhattacharya for allowing me to use his thesis material for these notes [4]

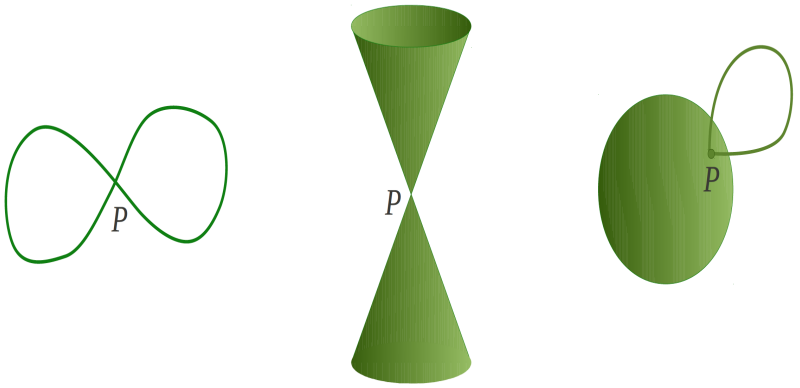


Figure 6.3: Topological spaces that are not manifolds. The spaces look locally Euclidean everywhere, except for the points marked as 'P'.

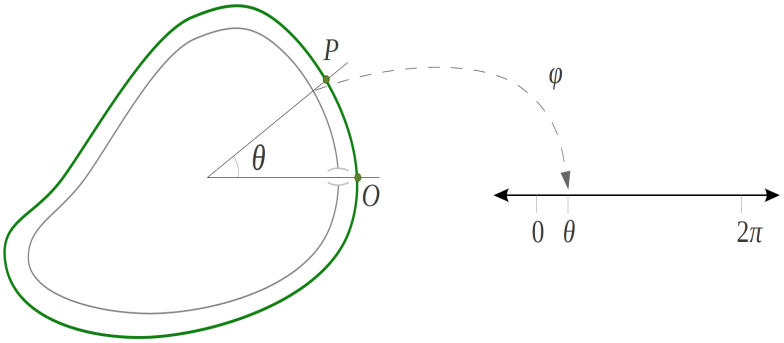


Figure 6.4: The map $\phi : (\mathbb{S}^1 - o) \longrightarrow (0, 2\pi)$ constitutes a coordinate chart. Note that $(\mathbb{S}^1 - o)$ is an open subset of $\mathbb{S}^1$.

since the neighborhood of every point on it resembles a line segment. Similarly a sphere or a torus are two-dimensional manifolds. A solid ball is a three-dimensional manifold with boundary. Figure 6.3 shows some simple topological spaces that are not manifolds. This is because at least at some point in the spaces there does not exist an open neighborhood that resembles an Euclidean space.

Next we proceed towards defining a coordinate chart on a manifold. There are many natural algebraic tools associated with the standard Euclidean space(e.g. its vector-space structure,natural definition of differentiation along the axes, etc.). The main purpose of defining a coordinate chart is to borrow those concepts to arbitrary manifolds.

Definition 48. (Coordinate Chart [5]). Given an open subset $\mathbb{U}$ of a d-dimensional manifold $\mathbb{M}$, and a continuous injective function $\phi : \mathbb{U} \longrightarrow \mathbb{R}^d$, we say $C = (\mathbb{U}, \phi)$ is a coordinate chart on $\mathbb{U}$. ϕ is a homeomorphism over its image.

Thus, the polar coordinate $\theta \in (0, 2\pi)$ used to describe points on a circle Figure 6.4 constitutes a coordinate chart. The map $\phi : (\mathbb{S}^1 - o) \longrightarrow (0, 2\pi)$ maps every points on the circle, except one (exclusion of which makes the pre-image of ϕ an open subset of $\mathbb{S}^1$ 1, and the image an open subset of $\mathbb{R}^1$), to the open interval $(0, 2\pi)$ on $\mathbb{R}$. Similarly, the familiar polar coordinate on a 2-sphere constitutes a coordinate chart. The open subset under consideration here is the entire surface of the sphere except the polar points and one longitudinal line, and the function ϕ maps every point on it to a point in $\mathbb{R}^2$, namely, (θ, γ) - the latitudinal and longitudinal angles.

The variable (θ in case of the circle, (θ, γ) in case of a sphere), natural to $\mathbf{R}^d$, can now be used to describe points on the pre-image of ϕ i.e. the open subset $\mathbb{U}$ (since ϕ. has a continuous inverse as well). These are called the coordinate variables of the chart $\mathbf{C}$. Very often, if $\mathbf{x} = \{x^1, x^2, \cdots, x^d\}$ are the coordinate variables corresponding to a given chart, one simply writes $\mathbf{x}$ to refer to points on$\mathbb{U}$ instead of writing $\phi(\mathbf{x})$.

Definition 49. (Atlas)[5] Figure 6.5. An atlas is a collection of coordinate charts $\{\mathbb{U}_\alpha, \phi_\alpha\}$ on a manifold $\mathbb{M}$, such

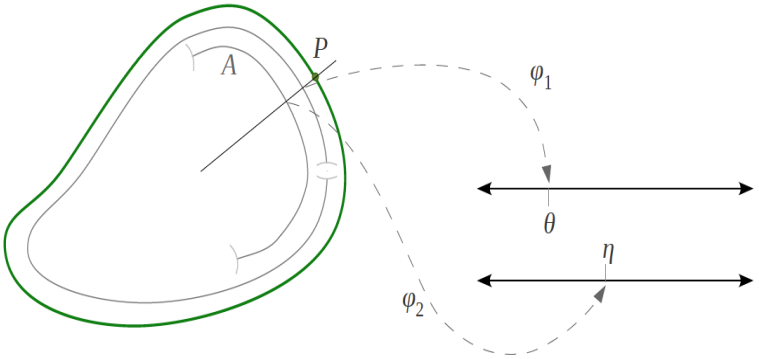


Figure 6.5: Two maps, $\phi_1 : (\mathbb{S}^1 - o) \longrightarrow \mathbb{R}$ and $\phi_2 : A \longrightarrow \mathbb{R}$ with A being an open subset of $\mathbb{S}^1$ constitutes an atlas on $\mathbb{S}^1$. This is because $(\mathbb{S}^1 - o) \cup A = \mathbb{S}^1$.

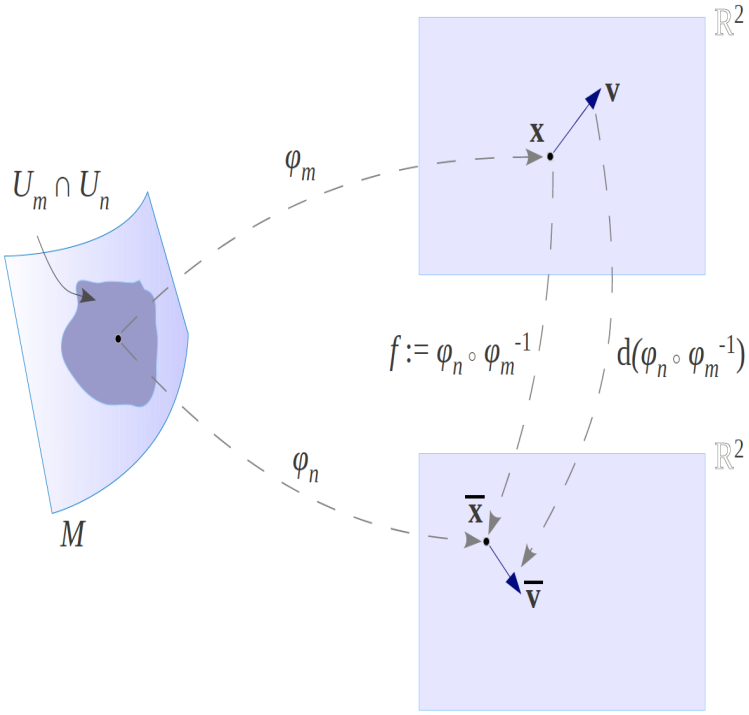


Figure 6.6: Chart Transition

that the union of the open subsets covers the entire of $\mathbb{M}$ (we say $\mathbb{U}_\alpha$ is an open covering of $\mathbb{M}$). That is $\mathbb{M} \subseteq \cup_\alpha \mathbb{U}_\alpha$

Going back to the example of the chart on a open subset of circle, we were unable to incorporate one single point on the circle for being described by the chart in Figure 6.4. However, with an atlas (Figure 6.5), we can have multiple separate charts, using which we can cover the entire circle.

Definition 50. (Chart Transition Figure 6.6). Consider two charts $\{\mathbb{U}_m, \phi_m\}$ and $\{\mathbb{U}_n, \phi_n\}$, for some open subsets $\mathbb{U}_m$ and $\mathbb{U}_n$ of a d-dimensional manifold $\mathbb{M}$ such that $\mathbb{U}_m \cap \mathbb{U}_n \neq \emptyset$. Then one can define the map $\phi_n \circ \phi_m^{-1} : \phi_m(\mathbb{U}_m \cap \mathbb{U}_n) \longrightarrow \phi_n(\mathbb{U}_m \cap \mathbb{U}_n)$. Note that both the pre-image and image of this map are subsets of $\mathbb{R}^d$. Such a map is called a chart transition or a coordinate transformation or simply a coordinate change.

A manifold is said to be differentiable if for every pair of charts $\{\mathbb{U}_m, \phi_m\}$ and $\{\mathbb{U}_n, \phi_n\}$ the transitions $\phi_n \circ \phi_m^{-1}$ are C^∞ differentiable functions.

6.2 Tangent space

We are familiar with the notion of a tangent to a curve or the tangent plane to a surface embedded in $\mathbb{R}^3$. The notion of such tangents is generalized to arbitrary manifolds by tangent space. Consider a chart $\{\mathbb{U}_m, \phi_m\}$. A point, $p \in \mathbb{U}_m$ is represented by its coordinates $\phi_m(p) = \mathbf{x} = [x^1, x^2, \cdots, x^d] \in \mathbb{R}^d$. Then one can imagine a vector centered at $\mathbf{x}$ and having components $[v^1, v^2, \cdots, v^d]$ along the directions of increasing values of $[x^1, x^2, \cdots, x^d]$ Figure 6.6.

Fundamental to the definition of such a vector is the notion of transformation of the coefficients, $[v^1, v^2, \cdots, v^d]$, under a chart transition. Consider the chart transition $\hat{x} = f(\mathbf{x})$ as discussed earlier. The vector $\mathbf{v}$ (sitting at $\mathbf{x}$) with components $[v^1, v^2, \cdots, v^d]$ in the un-barred coordinate variables will have certain components $[\hat{v}^1, \hat{v}^2, \cdots, \hat{v}^d]$ in the barred coordinates (sitting at $\hat{x}$) Figure 6.6, which is determined by some transformation rule that we are yet to define. However, whatever that definition be, the property that such a transformation rule on coefficient vectors attached to points, x or $\hat{x}$, must satisfy is invariance under a composition of forward and inverse transformation.

Contravariant Vectors: One such transformation rule for vectors that satisfies the above conditions is $\hat{v}^j \sum_{i=1}^d \frac{\partial f^j}{\partial x^i}|_{\mathbf{x}} v^i$. This can be easily seen by noting that $[\hat{v}] = \mathbf{J}_f[v]$ (where $\mathbf{J}_f$ is the Jacobian matrix of f at $\mathbf{x}$, and $[v]$ represents the coefficient vector as a column vector). Also, $[v] = \mathbf{J}_g[\hat{v}]$ and therefore $\mathbf{J}_g\mathbf{J}_f = \mathbf{I}$. Such vectors, the coefficients of which follow such transformation rules, are known as *contravariant* vectors (or simply, *vectors*). There are other types of vectors that transform differently, but satisfy the said properties under composition of transformations. One such example is that of covariant vectors, which we will not discuss in details in this section.

A systematic way of writing contravariant vectors is by choosing $\frac{\partial}{\partial x^i}$ as basis for such vectors (i.e.quantities to which we 'multiply' the said coefficients/components, v^i, and take sum to write the full vector). This is purely done to avoid writing the said transformation rule for vector components explicitly, and instead take advantage of the standard rule for transformation of partial derivatives from one set of variables to another. Thus, under this representation, one would write for a vector, $\mathbf{v} = \sum_{i=1}^d v^i \frac{\partial}{\partial x^i}$ (where as usual the v^i are the components in the specific chart), and asserts that this v is independent of the choice of coordinate chart. That is,

$$\mathbf{v} = \sum_{i=1}^{d} v^i \frac{\partial}{\partial x^i} \;=\; \sum_{i=1}^{d} \hat{v}^i \frac{\partial}{\partial \hat{x}^i}$$

Now, from the chain rule of partial derivatives,

$$\frac{\partial}{\partial x^i} = \sum_{i=1}^{d} \frac{\partial \hat{x}^j}{\partial x^i} \frac{\partial}{\partial \hat{x}^j},$$

then

$$\hat{v}^j = \sum_{i=1}^{d} \frac{\partial \hat{x}^j}{\partial x^i} v^i \tag{6.1}$$

Definition 51. (Tangent space). Given a coordinate chart $\{\mathbb{U}, \phi\}$ on a smooth manifold $\mathbb{M}$, the tangent space at a point p on it represented by the coordinate variable $\mathbf{x} \in \Omega = Img(\phi) \subset \mathbb{R}^d$ is a d-dimensional vector space,

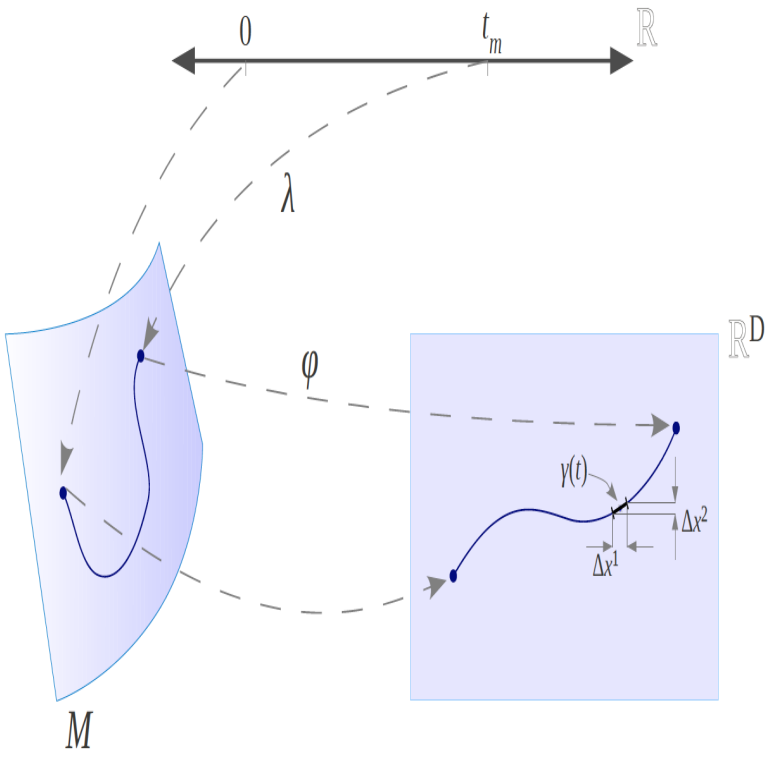


Figure 6.7: An infinitesimal element on a curve.

$T_p\mathbb{M} = T_p\Omega$, spanned by the basis $\{\frac{\partial}{\partial \hat{x}^1}, \frac{\partial}{\partial \hat{x}^2}, \cdots \frac{\partial}{\partial \hat{x}^d}\}$.

Remark. **Einstein's Summation Convention for Repeated Indices:** Purely for the convenience of writing, we would often drop the summation sign (capital sigma) inside expressions like $\sum_{i=1}^{d} \frac{\partial \hat{x}^j}{\partial x^i} v^i$ and whenever there is a repeated index (i in this case), we will assume the summation to be implied. Thus, for these expressions we will simply write $\frac{\partial \hat{x}^j}{\partial x^i}$.

6.3 Riemannian Geometry

We establish a structure on a smooth manifold that allows one to assign vectors in each tangent space a length (and an angle between vectors in the same tangent space). From this structure, one can then define a notion of length of a curve. Then we can look at shortest curves (which will be called *geodesics*). In this subsection, we will mostly be working with differentiable manifolds.

The definition of length of a curve on a d-dimensional manifold, $\mathbb{M}$, represented by $\lambda : [0, t] \longrightarrow \mathbb{M}$, is closely related with the definition of the length of an infinitesimal element on the curve. One can then integrate the infinitesimal lengths to obtain the total length of the curve. While the length of an infinitesimal element on a curve can have a variety of possible definitions (including ones based on arbitrary norms and higher order derivatives of the curve), the type of definition that we will be interested in is based on quadratic forms on the tangent spaces of the manifold. Such a definition of length arises naturally in many physical and practical applications, and forms the motivation for Riemannian metric.

For a given coordinate chart $\{\mathbb{U}, \phi\}$ with coordinate variables x^i , such that the curve λ lies entirely in $\mathbb{U}$, we define the curve in the given coordinates as $\gamma = \phi \circ \lambda : [0, t] \longrightarrow \mathbb{R}^d$ Figure 6.7. The domain of λ is called the parameter space of the curve. A small infinitesimal element on the curve between t and $t + \Delta t$ in the parameter space is then represented by $\dot{\gamma}\delta t = [\delta x^1, \delta x^2, \cdots, \delta x^d]$ (where x^i represent the change in the i^{th} coordinate variable

across the infinitesimal element in the chart. It is not difficult to note that $[\Delta \mathbf{x}] = [\Delta x^1, \Delta x^2, \cdots, \Delta x^d]$ behaves like coefficients of a contravariant vector (since in a different coordinate chart one would get $\Delta \hat{x}^j = \frac{\partial \hat{x}^j}{\partial x^i} \Delta x^i$ a consequence of elementary calculus).

Then we define the square of the 'length' of the element as

$$(\Delta s)^2 = \mathbf{G}_{ij} \Delta x^i \Delta x^j \tag{6.2}$$

g_{ij} hence represents elements of a matrix, $\mathbf{G}$ (which is specific to the given coordinate chart), such that the length is given by $[\Delta x]^T \mathbf{G} [\Delta x]$.

The condition that the definition of length must satisfy is that the length of an infinitesimal element should not change upon change of coordinates. Thus, if we are given another coordinate chart with coordinate variables $\hat{x}^i$, the following condition must hold,

$$(\Delta s)^2 = \mathbf{G}_{ij} \Delta x^i \Delta x^j = \hat{\mathbf{G}}_{pq} \Delta \hat{x}^p \Delta \hat{x}^q \tag{6.3}$$

Using the transformation rule for coefficients of contravariant vector, $\Delta x^i = \frac{\partial x^i}{\partial \hat{x}^p} \Delta x^p$ and substituting it in the middle term of the above equation, one gets the relationship between $\mathbf{G}_{ij}$ and $\hat{\mathbf{G}}_{pq}$

$$\hat{\mathbf{G}}_{pq} = \frac{\partial x^i}{\partial \hat{x}^p} \frac{\partial x^j}{\partial \hat{x}^q} \mathbf{G}_{ij} \tag{6.4}$$

The discussion so far indicates the definition of a bilinear scalar product on contravariant vectors (e.g. acting on two copies of $\mathbf{\Delta x} = \Delta x^i \frac{\partial}{\partial x^i}$ to give the square of length of the segment as prescribed by Equation 6.3. It is important to note that the value of this product for two contravariant vectors does not depend on the choice of the coordinate chart (this is achieved by the way we constructed the transformation rule Equation 6.4). Thus it is a product defined on the tangent spaces of the manifold itself.

Definition 52. (Riemannian Metric [14]). A Riemannian metric on a differentiable manifold, $\mathbb{M}$, is a symmetric bilinear scalar product on each tangent space, $T_p\mathbb{M}$, $p \in \mathbb{M}$, such that it varies smoothly with p. That is, it is a bilinear function $\mathbf{G}(p) : T_p\mathbb{M} \times T_p\mathbb{M} \longrightarrow \mathbb{R}_{\geq 0}$, that is symmetric in its two parameters, and it itself is smooth in p.

$\mathbf{G}$ is called the metric tensor, and the $\mathbf{G}_{ij}$, from our previous discussion, the matrix representation of the metric in a particular coordinate chart. The symmetry condition implies that the matrix representation is a symmetric matrix in any coordinate chart.

It is important to note that matrix representation of the metric in a particular coordinate chart simply constitutes a collections of $d(d+1)/2$ functions in the coordinate variables. That is,

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11}(\mathbf{x}) & \mathbf{G}_{12}(\mathbf{x}) & \cdots & \mathbf{G}_{1d}(\mathbf{x}) \\ \mathbf{G}_{21}(\mathbf{x}) & \mathbf{G}_{22}(\mathbf{x}) & \cdots & \mathbf{G}_{2d}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{d1}(\mathbf{x}) & \mathbf{G}_{d2}(\mathbf{x}) & \cdots & \mathbf{G}_{dd}(\mathbf{x}) \end{bmatrix} \tag{6.5}$$

with $\mathbf{G}_{ij}(\mathbf{x}) = \mathbf{G}_{ji}(\mathbf{x})$.

We use the following notation to write derivatives of the components of the matrix representation of the metric in a particular coordinate chart,

$$\mathbf{G}_{ij,k} = \frac{\partial \mathbf{G}_{ij}}{\partial x^k}$$

We will next describe a few quantities (specific to a particular coordinate chart) derived directly from the components of the matrix representation of the metric tensor on a particular coordinate system without detailing their immediate significance. We will hence use those quantities in order to state some results. The readers are directed to [14] for more detailed discussion on these quantities.

6.4 Geodesic

Consider the length of a curve $\gamma : [0, t] \longrightarrow \mathbb{R}^d$ on a Riemannian manifold (understand this in a coordinate sense i.e. $\gamma = \phi \circ \lambda$, where λ is the original curve and $\{\mathbb{U}, \phi\}$),

$$L(\gamma) = \int_0^t \sqrt{\mathbf{G}_{ij}(\gamma(\tau))\dot{\gamma}^i(\tau)\dot{\gamma}^j(\tau)}d\tau \tag{6.6}$$

where $\dot{\gamma}^i(\tau)$ is the i^{th} component of the tangent vector along γ at τ(in the chosen coordinates). Due to the way we defined $\mathbf{G}$ and its transformation, the length of a fixed curve on $\mathbb{M}$ is independent of the choice of the coordinate system.

If the start and end points of are constrained to two specific points (i.e $\gamma(0) = o \in \mathbb{R}^d$ and $\gamma(t) = d \in \mathbb{R}^d$), then one can consider the problem of minimizing $L(\gamma)$ over the different curves, that connect the two points. It can be shown that the γ that minimizes L, also minimizes the integral

$$E[\gamma] = \frac{1}{2}\int_0^t \mathbf{G}_{ij}(\gamma(\tau))\dot{\gamma}^i(\tau)\dot{\gamma}^j(\tau)d\tau. \tag{6.7}$$

The minimizing γ is called a geodesic curve or simply *geodesic*. This idea can also be formulated similar to what is done in classical mechanics where the equation of motion is a solution to the Euler-Lagrange equation for a given Lagrangian function. We now state a theorem along these lines

Theorem 6.1. *γ is geodesic if and only if it satisfies the Euler-Lagrange equations for the Lagrangian $\mathcal{L}$. The Lagrangian in coordinates is defined as:*

$$\mathcal{L}(\gamma, \dot{\gamma}) = \sqrt{\mathbf{G}_{ij}(\gamma)\dot{\gamma}^i\dot{\gamma}^j}. \tag{6.8}$$

Also, Euler-Lagrange equation in chart takes the form:

$$\frac{d}{d\tau}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^i}\right) - \frac{\partial \mathcal{L}}{\partial x^i} = 0 \tag{6.9}$$

The Euler-Lagrange equation can be reduced to a differential equation known as the *Geodesic Equation*. The differential equation (geodesic equation) can expressed as

$$\frac{d^2\gamma^i}{d\tau^2} + \Gamma^i_{jk}\frac{d\gamma^j}{d\tau}\frac{d\gamma^k}{d\tau} = 0, \tag{6.10}$$

where Γ^i_{jk} is the *Christoffel Symbol*, defined as:

$$\Gamma^i_{jk} = \frac{1}{2}\mathbf{G}_{im}\left(\mathbf{G}_{mk,l} + \mathbf{G}_{ml,k} - \mathbf{G}_{kl,m}\right) \tag{6.11}$$

Therefore, geodesics are solutions to Equation 6.10. It worth noting that, the solution to Equation 6.10 yields local optima not global optima. Thus, it is possible that more than one curve connecting two points satisfy the geodesic equation.

The length of the *shortest geodesic* (i.e. minimum over the lengths of all the possible geodesics between those points) between two points describe a distance function.

7 Lie Group and Lie algebra

Lie groups are mathematical objects named after the Norwegian mathematician Sophus Lie. These mathematical objects are extremely useful in studying symmetrical behaviors in geometry and differential equations. Since objects like rotation and translation operators[2] are also Lie groups, these ideas have been of great interest in robotics community too. The notes in section are primarily based on [21] and [8].

7.1 Lie Group

A Lie group is a group $\mathbb{G}$ which is also a smooth manifold and for which the group operations $\odot : \mathbb{G} \times \mathbb{G} \longrightarrow \mathbb{G}$ $(g \odot h \in \mathbb{G}\ \forall g, h \in \mathbb{G})$ and $g \longrightarrow g^{-1}\ g \in \mathbb{G}$ are smooth$(C^{\infty}(\mathbb{G}))$. A Lie group is *abelian* if the underlying group is commutative with respect to the $\odot$ operation.

Example 36. *(The Euclidean space under addition). The Euclidean space $\mathbb{R}^n$ under addition is an abelian Lie group.*

7.2 Matrix Lie group

A *matrix Lie group* $(\mathbb{G}, \odot)$ is a Lie group for which the underlying manifold is analytic and both the group operations are also analytic[3]. The Lie groups that are useful for robotics applications are mostly matrix Lie groups. Hence, we will be discussing only matrix Lie groups in this article. Specifically, we will be concentrating on subgroups of the space of invertible matrices or general linear group denoted as $\mathbf{GL}(n, \mathbb{R})$.

Example 37. *(The general linear group($\mathbf{GL}(n, \mathbb{R})$)). The group of all of $n \times n$ nonsingular real matrices is called the general linear group.*

Example 38. *(The special orthogonal group,$\mathbb{SO}(n)$). The special orthogonal group is a subgroup of the general linear group, defined as*

$$\mathbf{SO}(n) = \{\mathbf{R} \in \mathbf{GL}(n, \mathbb{R}) : \mathbf{R}^T\mathbf{R} = \mathbf{R}\mathbf{R}^T = \mathbf{I}, det(\mathbf{R}) = +1\}$$

The dimension of $\mathbf{SO}(n)$ as a manifold is $n(n-1)/2$. For $n = 3$, the group $\mathbf{SO}(3)$ is also referred to as the rotation group on $\mathbb{R}^3$.

Example 39. *The group of rigid transformations on $\mathbb{R}^3$ is defined as the set of mappings $r : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ of the form $r(x) = \mathbf{R}X + P$, where $\mathbb{R} \in \mathbf{SO}(3)$ and $P \in \mathbb{R}^3$. An element of $\mathbf{SE}(3)$ is written as $(P, \mathbf{R}) \in \mathbf{SE}(3)$. $\mathbf{SE}(3)$ can be identified with the space of 4×4 matrices of the form,*

$$r = \begin{bmatrix} \mathbf{R} & P \\ 0 & 1 \end{bmatrix}.$$

$\mathbf{SE}(3)$ is a Lie group of dimension 6.

7.3 Left and right translation

For any element $g \in \mathbb{G}$, we define *left translation* by g as the map $\mathbf{L}_g : \mathbb{G} \longrightarrow \mathbb{G}$ given by $\mathbf{L}_g(h) = gh$ for $h \in \mathbb{G}$. Also, right translation by g is defined as the map $\mathbf{R}_g(h) = hg$. Since $\mathbf{L}_g \circ \mathbf{L}_h = \mathbf{L}_{gh}$ and $\mathbf{R}_g \circ \mathbf{R}_h = \mathbf{R}_{gh}$, we have

[2] The right term is actions

[3] Taylor series of the operation map converges to the map

that $(\mathbf{L}_g)^{-1} = \mathbf{L}_{g^{-1}}$ and $(\mathbf{R}_g)^{-1} = \mathbf{R}_{g^{-1}}$. Thus, both $\mathbf{L}_g$ and $\mathbf{R}_g$ are diffeomorphisms of $\mathbb{G}$ for each g. Moreover, left and right translation commute: $\mathbf{L}_g \circ \mathbf{R}_h = \mathbf{R}_h \circ \mathbf{L}_g$. If the group is abelian then $\mathbf{L}_g = \mathbf{R}_g$.

7.4 Lie algebra

The tangent space at the identity element of a Lie group $\mathbb{G}$ with some additional algebraic structure in addition to the vector space is referred to as the Lie algebra of $\mathbb{G}$. Left(right) invariant vector fields are obtained by push forwarding elements in the Lie algebra with respect to the left(right) translations of all elements in the Lie group. Since every left (right) invariant vector fields can be constructed by push forwarding tangent vectors at the identity element, the space of all left(right) invariant vector fields are finite dimensional.

Apart from the vector space structure of a Lie algebra it is also endowed with a bilinear operator known as *Lie brackets* denoted as $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$, where $\mathfrak{g} \cong T_e\mathbb{G}$ is the Lie algebra associated with the $\mathbb{G}$. The result of the Lie bracket between $\eta_1 \in \mathfrak{g}$ and $\eta_2 \in \mathfrak{g}$ can be computed by taking the Lie bracket of their corresponding left invariant vector fields and evaluating the tangent vector of the resulting vector field at the identity element. Mathematically,

$$[\eta_1, \eta_2] = [X_{\eta_1}, X_{\eta_2}](e), \tag{7.1}$$

where, X_{η_1} and X_{η_2} are the left invariant vector fields associated with the η_1 and η_2 respectively. Since we are not discussing about Lie brackets in detail in this article, the reader may refer to [5, 15, 29] to learn about them.

Example 40. *(Lie algebra of $(\mathbb{R}^n, +)$). The groups identity element is origin ($e = 0$), $T_e\mathbb{R}^n \cong \mathbb{R}^n$. It is straightforward to see that the left invariant vector fields are constant vector fields of the form $X_v(x) = v$ for all $x \in \mathbb{R}^n$. Therefore, we can trivially compute $[v_1, v_2] = 0$.*

Example 41. *(The Lie algebra of $\mathbf{GL}(n, \mathbb{R})$ denoted as $\mathfrak{gl}(n, \mathbb{R})$). The Lie algebra is the set of all $n \times n$ real matrices, with the Lie bracket structure,*

$$[A, B] = AB - BA \quad A, B \in \mathfrak{gl}(n, \mathbb{R}).$$

Example 42. *(The Lie algebra of $\mathbb{SO}(3)$ denoted as $\mathfrak{so}(3)$). The Lie algebra may be identified with the 3×3 skew-symmetric matrices of the form,*

$$\hat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

where $\hat{\cdot}$[4] is the hat map, which maps the angular velocity vector ω to a skew symmetric matrix. The Lie algebra inherits the Lie bracket structure form $\mathfrak{gl}(n, \mathbb{R})$,

$$[\hat{\omega}_a, \hat{\omega}_b] = \hat{\omega}_a\hat{\omega}_b - \hat{\omega}_b\hat{\omega}_a, \quad \hat{\omega}_a, \hat{\omega}_b \in \mathfrak{so}(3).$$

One can easily show that,

$$[\hat{\omega}_a, \hat{\omega}_b] = \widehat{\omega_a \times \omega_b}, \quad \omega_a, \omega_b \in \mathbb{R}^3,$$

where $\times$ is the cross product operator between vectors.

Example 43. *(The Lie algebra of $\mathbb{SE}(3)$ denoted as $\mathfrak{se}(3)$). The Lie algebra of $\mathbb{SE}(3)$ can be identified with the*

[4] the hat map and vee map will be defined later

4×4 *skew-symmetric matrices of the form,*

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix}, \quad \omega, v \in \mathbb{R}^3.$$

Once again the Lie algebra inherits the Lie bracket structure form $\mathfrak{gl}(n, \mathbb{R})$, $[\hat{\xi}_a, \hat{\xi}_b] = \hat{\xi}_a\hat{\xi}_b - \hat{\xi}_b\hat{\xi}_a$. *Similar to* $\mathfrak{so}(3)$,

$$[\hat{\xi}_a, \hat{\xi}_b] = \begin{bmatrix} \widehat{(\omega_a \times \omega_b)} & \omega_a \times v_b - \omega_b \times v_a \\ 0 & 0 \end{bmatrix}$$

7.5 The Exponential and Logarithm Maps

Given a general matrix Lie group, elements sufficiently close to the identity are written as $g = \exp(\mathbf{X})$ for some $\mathbf{X} \in \mathfrak{g}$ (the Lie algebra of $\mathbb{G}$) with $\|\mathbf{X}\| \ll 1$. Here, the matrix Lie algebra $\mathbb{G}$ can be thought of as the set of all matrices $\{\mathbf{X}\}$ (not only ones for which $\|\mathbf{X}\| \ll 1$) such that the exponential of each $\mathbf{X}$ results in an element of $\mathbb{G}$.

Explicitly, in the context of matrix Lie groups (which are the ones used in robotics applications) the exponential map is simply the matrix exponential

$$\exp(\mathbf{X}) = \sum_{i=0}^{\infty} \frac{\mathbf{X}^i}{i!}. \tag{7.2}$$

The *exponential map* takes an element of the Lie algebra and produces an element of the Lie group. This is written as

$$\exp : \mathfrak{g} \longrightarrow \mathbb{G}.$$

As one would guess, the inverse map is called *logarithm map*:

$$\log : \mathbb{G} \longrightarrow \mathfrak{g}, \tag{7.3}$$

defined as

$$\log(g) = \log(e + (g - e)) = \sum_{i=0}^{\infty} (-1)^{i+1} \frac{(g-e)^i}{i}, \tag{7.4}$$

where e is identity element in the matrix Lie group which is the identity matrix. Additionally, although it is possible to exponentiate any element of a Lie algebra, the logarithm is only defined in a ball around the identity element of $\mathbb{G}$. In some cases, this ball extends over the whole of $\mathbb{G}$, or up to $\mathbb{G}$ minus a set of measure zero, but, in general, caution must be exercised in the application of the logarithm. It is noteworthy that, If $\mathbb{G}$ is compact it can be shown that the exponential map is surjective.

Example 44. *(The exponential and logarithmic map on $\mathbf{GL}(n,\mathbb{R})$). Let $\mathbf{a} \in \mathfrak{gl}(n,\mathbb{R})$, then*

$$\exp(\mathbf{a}) = \sum_{i=0}^{\infty} \frac{\mathbf{a}^i}{i!}.$$

If $\mathbf{A} \in \mathbf{GL}(n,\mathbb{R})$, then

$$\log(\mathbf{A}) = \sum_{i=0}^{\infty} (-1)^{i+1} \frac{(\mathbf{A}-\mathbf{I})^i}{i}, \tag{7.5}$$

which converges for all $\|\mathbf{A}-\mathbf{I}\| < 1$.

Example 45. *(The exponential and logarithmic map on $\mathbb{SO}(3)$). Let $\hat{\omega} \in \mathbb{SO}(3)$, then $\exp(\hat{\omega})$ corresponds to a rotation about the vector $\hat{\omega} \in \mathbb{R}^3$ by an angle $\|\omega\|$. An explicit formula is given by Rodrigues's formula [21]:*

$$\exp(\hat{\omega}) = \mathbf{I} + \frac{\hat{\omega}}{\|\omega\|} \sin\|\omega\| + \frac{\hat{\omega}^2}{\|\omega\|^2}(1-\cos\|\omega\|). \tag{7.6}$$

Also, if $\mathbf{R} \in \mathbb{SO}(3)$, then $\log \mathbf{R} = \hat{a} = \theta\hat{\omega}$, where $\theta \in \mathbb{R}$,

$$\cos\theta = \frac{trace(\mathbf{R}-1)}{2} \tag{7.7}$$

$$\hat{\omega} = \frac{1}{2\sin\theta}(\mathbf{R}-\mathbf{R}^T). \tag{7.8}$$

When $\mathbf{R} = \mathbf{I}, \theta = 2\pi k$ for any integer k then $\hat{\omega}$ can be chosen arbitrarily. Note that the log function is multi-valued since θ is not unique.

Example 46. *(The exponential and logarithmic map on $\mathbb{SE}(3)$). Again, the Lie algebra can be identified with 4×4 matrices of the form*

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix}, \omega, v \in \mathbb{R}^3 \tag{7.9}$$

with $[\hat{\xi}_a, \hat{\xi}_b] = \hat{\xi}_a\hat{\xi}_b - \hat{\xi}_b\hat{\xi}_a$. The exponential map is given by

$$\exp(\hat{\xi}) = \begin{bmatrix} \mathbf{I} & v \\ 0 & 1 \end{bmatrix}, \omega = 0 \tag{7.10}$$

and

$$\exp(\hat{\xi}) = \begin{bmatrix} e^{\hat{\omega}} & \mathbf{A}v \\ 0 & 1 \end{bmatrix}, \omega \neq 0, \tag{7.11}$$

where

$$\mathbf{A} = \mathbf{I} + \frac{\hat{\omega}}{\|\omega\|}(1-\cos\|\omega\|) + \frac{\hat{\omega}^2}{\|\omega\|^3}(\|\omega\| - \sin\omega). \tag{7.12}$$

The log function on $\mathbb{SE}(3)$ is given by,

$$\hat{\xi} = \log \begin{bmatrix} \mathbf{R} & p \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \hat{\omega} & \mathbf{A}^{-1}p \\ 0 & 0 \end{bmatrix}, \tag{7.13}$$

where $\hat{\omega} = \log \mathbf{R}$ *and*

$$\mathbf{A}^{-1} = \mathbf{I} - \frac{\hat{\omega}}{2} + \frac{2\sin\|\omega\| - \|\omega\|(1+\cos\|\omega\|)}{2\|\omega\|^2 \sin\omega}\hat{\omega}^2, \hat{\omega} \neq 0. \tag{7.14}$$

7.6 Hat($\hat{\cdot}$) and Vee($^\vee$) operators

Every element in the neighborhood of the identity of a connected matrix Lie group $\mathbb{G}$ can be described with the exponential parameterization

$$g = g(x^1, x^2, \cdots, x^n) = \exp\left(\sum_{i=1}^{n} x^i B_i\right) \tag{7.15}$$

where n is the dimension of the group and $\{B_i\}$ is a basis for its Lie algebra $\mathfrak{g}$ which is orthonormal with respect to a given inner product. For some Lie groups, the exponential parameterization extends over the whole group. Now, the "vee" of the Lie algebra is defined as

$$\exp\left(\sum_{i=1}^{n} x^i B_i\right)^\vee = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \\ \vdots \\ x^n \end{bmatrix}. \tag{7.16}$$

The hat operator is the inverse of vee operator.

7.7 Rigid body kinematics

The configuration of a rigid body with respect to some reference configuration is described by an element $g = (p, \mathbf{R}) \in \mathbb{SE}(3)$ If A is a fixed coordinate frame and B a frame attached to the rigid body, then we write $g_{ab} = (p_{ab}, \mathbf{R}_{ab} \in \mathbb{SE}(3))$ to denote the configuration of B with respect to A. p_{ab} represents the location of the origin of the B frame and $\mathbf{R}_{ab} \in \mathbb{SO}(3)$ its orientation. The group operation on $\mathbb{SE}(3)$ allows us to determine the configuration of a frame C relative to A via an intermediate frame B:

$$g_{ac} = g_{ab} \cdot g_{bc} = (p_{ab} + \mathbf{R}_{ab}P_{bc}, \mathbf{R}_{ab}\mathbf{R}_{bc}).$$

Therefore, any rigid body motion can be represented as curve on $\mathbb{SE}(3)$. If we represent $g \in \mathbb{SE}(3)$ as a 4×4 homogeneous matrix,

$$g = \begin{bmatrix} \mathbf{R} & p \\ 0 & 1 \end{bmatrix}$$

then the group operation is given by matrix multiplication and we may regard $\mathbb{SE}(3)$ as a subgroup of the general linear group, $\mathbf{GL}(4, \mathbb{R})$.

The configuration $g_{ab} \in \mathbb{SE}(3)$ can also be interpreted as a mapping from the coordinates of a point written relative to the B frame into the coordinates of the same point written relative to the A frame. Formally, this defines an action of $\mathbb{SE}(3)$ on $\mathbb{R}^3$ given by $\Phi_g(p) = p + \mathbf{R}q$ In homogeneous coordinates this action can be written as

$$\begin{bmatrix} q_a \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{ab} & p_{ab} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} q_b \\ 1 \end{bmatrix}$$

It follows from associativity of matrix multiplication that this actually defines an action of $\mathbb{SE}(3)$ on $\mathbb{R}^3$.

The action of $\mathbb{SE}(3)$ on vectors describes how the velocity of a point is mapped from one coordinate frame to another. Formally, we represent the velocity of a point as an element of $T_x\mathbb{R}^3$ and the action of $\mathbb{SE}(3)$ on tangent vectors (velocities) is the lifted action of $\mathbb{SE}(3)$ on $\mathbb{R}^3$. The lifted action of $\mathbb{SE}(3)$ on $T\mathbb{R}^3$ is given by $\Phi_{g*}(v_q) = (g(p), \mathbf{R}v_q)$ where $g(p)$ denotes the action of g on the point q. In homogeneous coordinates, the tangent space (velocity) portion of the action can be written as

$$\begin{bmatrix} v_a \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{ab} & p_{ab} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_b \\ 0 \end{bmatrix}$$

Since $\mathbb{SE}(3)$ is a Lie group, the exponential map can be used to map elements of the Lie algebra into the group. In homogeneous coordinates, the Lie algebra of $\mathbb{SE}(3)$ is a Lie subalgebra of $\mathfrak{gl}(4, \mathbb{R})$ consisting of matrices of the form

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix} \quad \hat{\omega} \in \mathfrak{so}(3), \ v \in \mathbb{R}^3$$

with the Lie bracket given by the matrix commutator. We call an element of the Lie algebra $\mathfrak{se}(3)$ a twist. A twist can be interpreted geometrically using the theory of screws. Consider the motion generated by simultaneously rotating and translating about an axis in the direction $\omega \in \mathbb{R}^3$ going through a point $q \in \mathbb{R}^3$. Let h represent the ratio of translational motion to rotational motion. If h is finite, then the resulting rigid motion is the exponential of the twist $\hat{\xi} \in \mathfrak{se}(3)$ given by

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & -q \times \omega + h\omega \\ 0 & 0 \end{bmatrix}$$

The one-parameter subgroup $\exp(\hat{\xi}\theta)$ generated by this twist corresponds to a rotation about an axis followed by translation along that same axis. Thus the exponential of a twist generates a screw motion.

For $\mathbb{SO}(3)$ it can be shown that the exponential map is actually surjective and hence any rigid transformation can be written as the exponential of some twist.

Let $q \in \mathbb{R}^3$ be a point attached to a rigid body and let $g_{ab} \in \mathbb{SE}(3)$ describe the trajectory of a frame B attached to the rigid body relative to a fixed reference frame A. In homogeneous coordinates, the trajectory of the point q as a function of time can be written as

$$q_a(t) = g_{ab} q_b$$

where q_a and q_b are the coordinates of the point relative to the A and B frames. The velocity of the point, relative

to the A frame, is given by

$$v_a(t) = \dot{q}_a(t) = \dot{g}_{ab}(t)q_b,$$

where we have used the fact that q_b is constant since the point is fixed in the body frame. Thus, $\hat{g}_{ab}(t) \in T_g\mathbb{SE}(3)$ can be viewed as a mapping between the body coordinates of a point and the spatial velocity of that same point. A more appealing representation of velocity is one which does not require switching between coordinate frames. That is, suppose we wish to find the relationship between the coordinates of a point and its velocity, when both quantities are specified with respect to a single frame. We can accomplish this by transforming either the coordinates of the point or the coordinates of velocity to the appropriate frame. For example, if we are given the coordinates of the point q with respect to the spatial frame A, then the velocity of q with respect to A is given by

$$v_a = \dot{g}q_b = (\dot{g}g^{-1})q_a$$

This is precisely the spatial velocity that is defined by using right translation to pull back the velocity $\dot{g} \in T_g\mathbb{SE}(3)$ to $T_e\mathbb{SE}(3)$. A similar argument shows that the body velocity, $g^{-1}\dot{g}$ can be viewed as a map from the body coordinates of a point to the body velocity of that point. The body and spatial velocities are physically interpreted as the instantaneous translational and rotational velocity written relative to the body or spatial frame, respectively.

8 Topology

Motivation: At the coarsest level, spacetime is a set. But, a set is not enough to talk about continuity of maps, which is required for classical physics notions such as trajectory of a particle. We do not want jumps such as a particle disappearing at some point on its trajectory and appearing somewhere. So we require continuity of maps. There could be many structures that allow us to talk about continuity, e.g., distance measure. But we need to be very minimal and very economic in order not to introduce undue assumptions. So we are interested in the weakest structure that can be established on a set which allows a good definition of continuity of maps. Mathematicians know that the weakest such structure is topology. This is the reason for studying topological spaces.

8.1 Topological Spaces

Definition 53. Let M be a set and $\mathcal{P}(M)$ be the power set of M, i.e., the set of all subsets of M.
A set $\mathcal{O} \subseteq \mathcal{P}(M)$ is called a **topology**, if it satisfies the following:

(i) $\emptyset \in \mathcal{O}$, $M \in \mathcal{O}$

(ii) $U \in \mathcal{O}, \quad V \in \mathcal{O} \implies U \cap V \in \mathcal{O}$

(iii) $U_\alpha \in \mathcal{O}, \quad \alpha \in \mathcal{A}$ ($\mathcal{A}$ is an index set $\implies \left(\bigcup_{\alpha\in\mathcal{A}} U_\alpha\right) \in \mathcal{O}$

Terminology:

1. the tuple $(M, \mathcal{O})$ is a **topological space**.
2. $\mathcal{U} \in M$ is an **open set** if $\mathcal{U} \in \mathcal{O}$.
3. $\mathcal{U} \in M$ is a **closed set** if $M \setminus \mathcal{U} \in \mathcal{O}$.

Definition 54. $(M, \mathcal{O})$, where $\mathcal{O} = \{\emptyset, M\}$ is called the **chaotic topology**.
Definition 55. $(M, \mathcal{O})$, where $\mathcal{O} = \mathcal{P}(M)$ is called the **discrete topology**.
Definition 56. A **soft ball** at the point p in $\mathbb{R}^d$ is the set

$$\mathcal{B}_r(p) := \left\{ (q_1, q_2, ..., q_d) \quad | \quad \sum_{i=1}^{d} (q_i - p_i)^2 < r^2 \right\} \text{ where } r \in \mathbb{R}^+ \tag{8.1}$$

Definition 57. $(\mathbb{R}^d, \mathcal{O}_{std})$ is the **standard topology**, provided that $U \in \mathcal{O}_{std}$ iff
$\forall p \in U, \exists r \in \mathbb{R}^+ : \mathcal{B}_r(p) \subseteq U$

Proof. $\emptyset \in \mathcal{O}_{std}$ since $\forall\, p \in \emptyset$, $\exists r \in \mathbb{R}^+$: $\mathcal{B}_r(p) \subseteq \emptyset$ (i.e. satisfied "vacuously")
$\mathbb{R}^d \in \mathcal{O}_{std}$ since $\forall p \in \mathbb{R}^d$, $\exists r = 1 \in \mathbb{R}^+$: $\mathcal{B}_r(p) \subseteq \mathbb{R}^d$

Suppose $U, V \in \mathcal{O}_{std}$. Let $p \in U \cap V \implies \exists\, r_1, r_2 \in \mathbb{R}^+$ s.t. $\quad \mathcal{B}_{r_1}(p) \subseteq U, \quad \mathcal{B}_{r_2}(p) \subseteq V$.
Let $r = \min\{r_1, r_2\} \implies \mathcal{B}_r(p) \subseteq U$ and $\mathcal{B}_r(p) \subseteq V \implies \mathcal{B}_r(p) \subseteq U \cap V \implies U \cap V \in \mathcal{O}_{std}$.

Suppose, $U_\alpha \in \mathcal{O}_{std}, \forall\, \alpha \in \mathcal{A}$. Let $p \in \bigcup_{\alpha\in\mathcal{A}} U_\alpha \implies \exists \alpha \in \mathcal{A} : p \in U_\alpha$
$\implies \exists\, r \in \mathbb{R}^+ : \mathcal{B}_r(p) \subseteq U_\alpha \subseteq \bigcup_{\alpha\in\mathcal{A}} U_\alpha \implies \bigcup_{\alpha\in\mathcal{A}} U_\alpha \in \mathcal{O}_{std}$. □

8.2 Continuous maps

A map f, $f : M \longrightarrow N$, connects each element of a set M (domain set) to an element of a set N (target set).

Terminology:

1. If f maps $m \in M$ to $n \in N$, then we may say $f(m) = n$, or m maps to n, or $m \mapsto f(m)$ or $m \mapsto n$.
2. If $V \subseteq N, \text{preim}_f(V) := \{m \in M | f(m) \in V\}$
3. If $\forall n \in N, \exists m \in M : n = f(m)$, then f is **surjective**. Or, $f : M \twoheadrightarrow N$.
4. If $m_1, m_2 \in M, m_1 \neq m_2 \implies f(m_1) \neq f(m_2)$, then f is **injective**. Or, $f : M \hookrightarrow N$.

Definition 58. Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces. A map $f : M \longrightarrow N$ is called **continuous** w.r.t. $\mathcal{O}_M$ and $\mathcal{O}_N$ if $V \in \mathcal{O}_N \implies (\text{preim}_f(V)) \in \mathcal{O}_M$.

***Mnemonic**: A map is continuous iff the preimages of all open sets are open sets.*

8.3 Composition of continuous maps

Definition 59. If $f : M \longrightarrow N$ and $g : N \longrightarrow P$, then

$$g \circ f : M \longrightarrow P \text{ such that } m \mapsto (g \circ f)(m) := g(f(m))$$

Theorem 8.1. *If $f : M \longrightarrow N$ is continuous w.r.t. $\mathcal{O}_M$ and $\mathcal{O}_N$ and $g : N \longrightarrow P$ is continuous w.r.t. $\mathcal{O}_N$ and $\mathcal{O}_P$, then $g \circ f : M \longrightarrow P$ is continuous w.r.t. $\mathcal{O}_M$ and $\mathcal{O}_P$.*

Proof. Let $W \in \mathcal{O}_P$.

$$\begin{aligned}
\text{preim}_{g\circ f}(W) &= \{m \in M | g(f(m)) \in W\} && \because (g \circ f)(m) = g(f(m)) \\
&= \{m \in M | f(m) \in \text{preim}_g(W)\} && \text{preim}_g(W) \in \mathcal{O}_N \because g \text{ is continuous} \\
&= \text{preim}_f(\text{preim}_g(W)) && \in \mathcal{O}_M \because f \text{ is continuous} \\
&\implies g \circ f \text{ is continuous}
\end{aligned}$$

□

8.4 Inheriting a topology

Given a topological space $(M, \mathcal{O}_M)$, one way of inheriting a topology from it is the subspace topology.

Theorem 8.2. *If $(M, \mathcal{O}_M)$ is a topological space and $S \subseteq M$, then the set $\mathcal{O}|_S \subseteq \mathcal{P}(S)$ such that $\mathcal{O}|_S := \{S \cap U | U \in \mathcal{O}_M\}$ is a topology. $\mathcal{O}|_S$ is called the **subspace topology** inherited from $\mathcal{O}_M$.*

Proof.

1. $\emptyset, S \in \mathcal{O}|_S \because \emptyset = S \cap \emptyset, S = S \cap M$.
2. $S_1, S_2 \in \mathcal{O}|_S \implies \exists U_1, U_2 \in \mathcal{O}_M : S_1 = S \cap U_1, S_2 = S \cap U_2 \implies U_1 \cap U_2 \in \mathcal{O}_M$
   $\implies S \cap (U_1 \cap U_2) \in \mathcal{O}|_S \implies (S \cap U_1) \cap (S \cap U_2) \in \mathcal{O}|_S \implies S_1 \cap S_2 \in \mathcal{O}|_S$.
3. Let $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is an index set. Then $S_\alpha \in \mathcal{O}|_S \implies \exists U_\alpha \in \mathcal{O}_M : S_\alpha = S \cap U_\alpha$.
   Further, let $\mathcal{U} = \left(\bigcup_{\alpha\in\mathcal{A}} U_\alpha\right)$. Therefore, $\mathcal{U} \in \mathcal{O}_M$.
   Now, $\left(\bigcup_{\alpha\in\mathcal{A}} S_\alpha\right) = \left(\bigcup_{\alpha\in\mathcal{A}} (S \cap U_\alpha)\right) = S \cap \left(\bigcup_{\alpha\in\mathcal{A}} U_\alpha\right) = S \cap \mathcal{U} \implies \left(\bigcup_{\alpha\in\mathcal{A}} S_\alpha\right) \in \mathcal{O}|_S$.

□

Theorem 8.3. *If $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ are topological spaces, and $f : M \longrightarrow N$ is continuous w.r.t $\mathcal{O}_M$ and $\mathcal{O}_N$, then the restriction of f to $S \subseteq M$, $f|_S : S \longrightarrow N$ s.t. $f|_S(s \in S) = f(s)$, is continuous w.r.t $\mathcal{O}|_S$ and $\mathcal{O}_N$.*

Proof. Let $V \in \mathcal{O}_N$. Then, $\text{preim}_f(V) \in \mathcal{O}_M$.
Now $\text{preim}_{f|_S}(V) = S \cap \text{preim}_f(V) \implies \text{preim}_{f|_S}(V) \in \mathcal{O}|_S \implies f|_S$ is continuous. □

References

[1] Sheldon Jay Axler, *Linear algebra done right*, Undergraduate Texts in Mathematics, Springer, New York, 1997.

[2] Richard Bellman et al., *The theory of dynamic programming*, Bulletin of the American Mathematical Society **60** (1954), no. 6, 503–515.

[3] Dimitri P. Bertsekas, *Dynamic programming and optimal control*, 2nd ed., Athena Scientific, 2000.

[4] Subhrajit Bhattacharya, *Topological and geometric techniques in graph search-based robot planning*, University of Pennsylvania, 2012.

[5] William Munger Boothby, *An introduction to differentiable manifolds and Riemannian geometry; 2nd ed.*, Pure Appl. Math., Academic Press, Orlando, FL, 1986.

[6] Francesco Bullo and Andrew D. Lewis, *Geometric control of mechanical systems*, Springer, 2004.

[7] Chi-Tsong Chen, *Linear system theory and design*, Oxford University Press, Inc., 1998.

[8] Gregory S Chirikjian, *Stochastic models, information theory, and lie groups, volume 2*, Springer, 2012.

[9] Sadri Hassani, *Mathematical physics: A modern introduction to its foundations*, 2nd ed., Springer, 2013.

[10] João P. Hespanha, *Linear systems theory*, Princeton Press, Princeton, New Jersey, 2018. ISBN13: 9780691179575.

[11] T.W. Judson, *Abstract algebra: Theory and applications*, The Prindle, Weber & Schmidt Series in Advanced Mathematics, PWS Publishing Company, 1994.

[12] T. Kailath, *Linear systems*, Information and System Sciences Series, Prentice-Hall, 1980.

[13] H. K. Khalil, *Nonlinear systems*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1996.

[14] John M Lee, *Riemannian manifolds: an introduction to curvature*, Vol. 176, Springer Science & Business Media, 2006.

[15] Jean Levine, *Analysis and control of nonlinear systems: A flatness-based approach*, Springer Science & Business Media, 2009.

[16] F.L. Lewis, V.L. Syrmos, and V.L. Syrmos, *Optimal control*, A Wiley-interscience publication, Wiley, 1995.

[17] Daniel Liberzon, *Calculus of variations and optimal control theory: A concise introduction*, Princeton University Press, Princeton, NJ, USA, 2011.

[18] J.E. Marsden, M.J. Hoffman, and U.M.J. Hoffman, *Elementary classical analysis*, W. H. Freeman, 1993.

[19] Phillipe Martin, Richard M Murray, and Pierre Rouchon, *Flat systems, equivalence and trajectory generation* (2003).

[20] J.R. Munkres, *Topology*, Featured Titles for Topology Series, Prentice Hall, Incorporated, 2000.

[21] Richard M Murray, *A mathematical introduction to robotic manipulation*, CRC press, 2017.

[22] W. Rudin, *Principles of mathematical analysis*, International series in pure and applied mathematics, McGraw-Hill, 1976.

[23] S. Sastry, *Nonlinear systems: Analysis, stability, and control*, Interdisciplinary Applied Mathematics, Springer New York, 1999.

[24] J. E. Slotine and W. Li, *Applied nonlinear control*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[25] Eduardo D Sontag, *Mathematical control theory: deterministic finite dimensional systems*, Vol. 6, Springer Science & Business Media, 2013.

[26] Gilbert Strang, *Linear algebra and its applications*, Thomson, Brooks/Cole, Belmont, CA, 2006.

[27] T. Tao, *Analysis*, Texts and Readings in Mathematics, Hindustan Book Agency, 2006.

[28] Russ Tedrake, *Underactuated robotics: Algorithms for walking, running, swimming, flying, and manipulation (course notes for mit 6.832)*, 2019.

[29] M. Vidyasagar, *Nonlinear systems analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1993.