Sub-Riemannian Problem on Lie Group of Motions of Pseudo Euclidean Plane

(Optimal Control and Optimality Analysis)

by

PE 113007 Yasir Awais Butt

A thesis submitted to the Electrical Engineering Department in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING Faculty of Engineering, Mohammad Ali Jinnah University, Islamabad, Pakistan.

June 2015

Dedicated to the memory of my beloved father.

CREDITS

Past three years have revealed upon me that PhD is the outcome of sheer hard work and infrangible dedication of not only the PhD scholar but many others connected to the scholar whose support, forbearance and commitment help the scholar walk through this long and laborious journey. Like every PhD candidate, I have gone through a process which was both exceedingly rewarding and terribly painful. At times I have stumbled, faltered and wavered and it is only due to the endless support of people around me that I always gathered myself and set on the journey again. For this sincere patronage I am deeply indebted to everyone around me. Unfortunately, neither would I remember everyone who lent a helping hand, nor it is possible to collect all the names here. Therefore, I offer my deepest gratitude to following for their endless support throughout my quest for knowledge and beg pardon from those countably infinite people whom I forgot:

- ▷ Allah Almighty who gave me resources, environment, strength and intellect to undertake this work.

- ▷ My parents and family who have provided me motivation and relentless support my entire life. My pressing commitment these past years has strained them enormously. Yet, they have been accommodating, bearing me out in all possible ways. My children have missed their playmate and my wife has missed a listening ear. Many family events were missed and mishandled and most were forgotten. For all this and everything else I am grateful to all my family.

- ▷ My supervisor, Dr. Aamer Iqbal Bhatti, whose unwavering trust in my abilities and insightful guidance enabled me to reach this point. He has been a trustworthy friend and a meticulous mentor befitting the very definition of a teacher. He has lent me practically unbounded administrative, academic, financial and moral support for which I am certainly most grateful.

▷ My co-supervisor Dr. Yuri L. Sachkov, Program System Institute, Pareslavl Zalessky, Russia. Accepting a student from Pakistan without any prior acquaintance and with humble background in Mathematical Control Theory bears testimony to his quest for knowledge and enthusiasm to spread education. He posed the problem and provided me with tremendous support, insightful discussions, enlightening explanations and thorough reviews for which I am deeply indebted. I am deeply impressed not only with his knowledge, but with his values, his personality, his kindness, his thoroughness, his time management, his ability of multitasking and his amazing reading skill that can find smallest and stupidest errors. I will always be in debt to him for his efforts to make me a better researcher and a better human being.

▷ My special thanks to Prof. A. Yu. Popov for his help with the proof of upper bound of first conjugate time.

▷ My colleagues at CASPR group who have patiently listened to my thoughts and have given me useful recommendations for my work. On most occasions my work has made little sense to the group members due to my haste in covering the material and delivering the presentations, yet, they listened hard and came up with ideas that helped me shape my work.

▷ I am particularly grateful to CASPR group members Imran Khan, Safdar Hussain, Ali Arshad, Ahmed Yar, Zeeshan Babar, Rizwan Azam for lending a helping hand in times of crisis.

▷ The members of examination committee for synopsis and thesis defense were indeed very helpful in reviewing the documents. Their guidance proved to be vital in properly shaping this document.

▷ We also wish to thank Prof. A. Yu. Popov for his help in the proof of Lemma 9.3.3.

▷ In the end, my special thanks to thousands of authors whose papers, web pages, codes, proofs and explanations helped build my understanding of the subject and carryout this work. Unfortunately, I cannot cite all of the work that I consulted and give due credit to the authors mostly because the work is not exclusively part of this thesis and partly because I have even forgotten the actual sources. Yet, their work stays with me and I pledge to pass it on as a repayment of my debt.

LIST OF PUBLICATIONS

- ▷ Yasir Awais Butt, Yuri L. Sachkov, Aamer Iqbal Bhatti, "Maxwell Strata and Conjugate Points in the Sub-Riemannian Problem on the Lie Group SH(2)", Submitted to SIAM Journal of Control and Optimization
- ▷ Yasir Awais Butt, Aamer Iqbal Bhatti, Yuri L. Sachkov, "Integrability by Quadratures in Optimal Control of a unicycle on a hyperbolic plane", Accepted for presentation in American Control Conference, Chicago, Illionis, 1–3 Jul 2015
- ▷ Yasir Awais Butt, Yuri L. Sachkov, Aamer Iqbal Bhatti, "Extremal Trajectories and Maxwell Strata in Sub-Riemannian Problem on Group of Motions of Pseudo-Euclidean Plane", Journal of Dynamical and Control Systems: Volume 20, Issue 3 (2014), Pages 341–364
- ▷ Yasir Awais Butt, Yuri L. Sachkov, Aamer Iqbal Bhatti, "Maxwell Strata and Conjugate Points in Sub-Riemannian Problem on Group SH(2)", International Youth Conference Geometry and Control, Moscow, April 14–18, 2014
- ▷ Yasir Awais Butt, Yuri L. Sachkov, Aamer Iqbal Bhatti, "Extremal Trajectories and Maxwell Points in Sub-Riemannian Problem on the Group SH(2)", International Conference on Mathematical Control Theory and Mechanics, Suzdal, Russia, 5–9 Jul 2013

ABSTRACT

The sub-Riemannian problem on group of motions of pseudo Euclidean plane is considered. From engineering perspective, this is the optimal control problem of a unicycle moving on a hyperbolic plane (plane with constant negative curvature). The dynamical system comprises real analytic left invariant vector field with 2-dimensional linear control vector - the control variables being the translational and the angular velocity of the unicycle. The sub-Riemannian optimal control problem seeks to determine optimal control input and the corresponding optimal trajectory between the given initial and terminal states such that the sub-Riemannian length on this trajectory is minimized.

At the onset we prove the controllability of the control distribution. We define the left invariant Hamiltonian for the system under consideration and apply the Pontryagin's Maximum Principle (PMP). We prove that the extremal trajectories in the abnormal case are not strictly abnormal and the corresponding control vector is a constant i.e., identically zero. We then consider the normal Hamiltonian system. Through suitable coordinate transformation, we prove that the vertical subsystem is a double covering of a mathematical pendulum. This fact allows us to introduce Jacobi elliptic functions for integration of the nonlinear state equations that would otherwise be analytically intractable. Using specific elliptic coordinates, we calculate the extremal trajectories parametrized by Jacobi elliptic functions corresponding to various energy levels of the pendulum describing the vertical subsystem. We gain further insight into the nature of extremal trajectories through simulation and qualitative analysis.

As PMP gives only first order necessary optimality conditions, the extremal trajectories resulting from the integration of normal Hamiltonian system are candidate optimal only. Hence, second order optimality conditions are applied to eliminate the trajectories that cease to be optimal at certain point. It is known that a normal extremal trajectory ceases to be optimal either because it meets another candidate optimal trajectory at a point called Maxwell point where both have equal sub-Riemannian length, or because there exists a point called a conjugate point where a family of extremal trajectories has an envelope. We find surfaces in the state space M containing all Maxwell points and obtain a description of the Maxwell points in terms of roots of function forming these surfaces. Once the Maxwell sets are calculated, an upper bound on the cut time (the time at which an extremal trajectory loses global optimality) is obtained.

As argued, an extremal trajectory can lose optimality due to the existence of conjugate points. Conjugate points are the critical points of the exponential mapping and are found as roots of the Jacobian of the exponential mapping. The time at which the first conjugate point exists is called the first conjugate time. We conclude that the first conjugate point occurs later than the first Maxwell point and therefore the upper bound on cut time is given by the first Maxwell time. In the end, we present 3-dimensional plots of some important objects in sub-Riemannian problem on Lie group SH(2).

Contents

List of Figures

Nomenclature

$[f, g]$	Lie bracket of vector fields f and g
λ	Costate variables
$\mathcal{L}$	Lie algebra
Exp	Exponential mapping
SH(2)	Lie group SH(2)
sn, cn, dn	Jacobi elliptic function sn, cn, dn
$\overrightarrow{h}$	Hamiltonian vector field
H	Hamiltonian function
h_u^ν	Control dependent Hamiltonian function used in PMP
T^*M	Cotangent bundle
t_1^{conj}	First conjugate time along a geodesic
t_1^{Max}	First Maxwell time along a geodesic
t_{cut}	Cut time i.e., time of loss of global optimality
T_x^*M	Cotangent space at point x of the manifold M
T_xM	Tangent space at point x of the manifold M
TM	Tangent bundle to the manifold M
f, g	Poisson bracket of functions f and g
Conj	Set of Conjugatel points
Max	Set of Maxwell points
PMP	Pontryagin's Maximum Principle
Vec(M)	Set of vector fields on the manifold M

Chapter 1

Introduction

> *I feel engulfed in the infinite immensity of spaces whereof I know nothing, and which know nothing of me, I am terrified. The eternal silence of these infinite spaces alarms me.*
>
> *Look somewhere else for someone who can follow you in your researches about numbers. For my part, I confess that they are far beyond me, and I am competent only to admire them.*
>
> (Blaise Pascal)

Having spent couple of years working on Geometric Control Theory and Sub-Riemannian geometry, the quotes above from Blaise Pascal express my feelings in the most appropriate manner. I must confess that I have stumbled to grasp and understand the subject as best as I could, yet, it was way more deeper and richer than could be covered in the given time. The sub-Riemannian problem on the Lie group SH(2) is an optimal control problem of hypothetical unicycle on a hyperbolic plane i.e., a plane with constant negative curvature. Like a standard optimal control problem it comprises two essential parts i.e., computation of optimal control function and finding the path on the configuration manifold that minimizes some cost functional. We employ tools from the geometric control theory and calculus of variations to compute the open loop optimal control. A detailed optimality analysis based on Maxwell points and conjugate points is carried out to eliminate the non-optimal extremal trajectories. Based on the optimality analysis, we are able to state an exact bound on the cut time i.e., the time when an extremal trajectory loses global optimality.

Being primarily an electronics engineer and with humble background in mathematics, much of the hardship I faced during this work is attributed to the lack of appropriate skill set to understand and admire the work of others and contribute something that could push the boundaries of the existing knowledge on this subject. It turned out that more than myself pushing knowledge boundaries outward, I was rather pushed inwards on most occasions. At the time of writing this thesis, I face yet another dilemma i.e., whether to give the

thesis a mathematician's perspective or an engineer's perspective. For someone initiated in mathematics, document could be written with little to no background buildup stating the theoretical results without giving much attention to intuitive or physical explanations. On the contrary lot of background knowledge with intuitive explanations of theoretical results, detailed explanation of terminologies, connection between various notions of optimal control etc. is needed to make it more accessible to the engineers. The work actually pertains to mathematical control theory and has a much less applied flavor, yet I fell in favor of the engineer's perspective. There are two compelling reasons for it i.e., first all along I tried to develop intuitive reasoning of various notions myself and I hope there would be others who would like it; and second I expect the understanding I developed over these couple of years to fade with time and only a detailed document could preserve that. Consequently, I have tried to gather essential definitions of various terminologies and various notions related to optimal control theory and sub-Riemannian geometry. Hence, a mathematician may find it laborious and overstretched and perhaps an engineer would find it appropriate. This however does not change the very persona of the research work. I do try to give intuitive explanations of various results yet the results are primarily theoretical in character and must be treated as such. Hence, one must not expect usual results that are found in a research work that pertains to control of physical systems i.e., there would be no closed loop simulation plots, there would be no experimental results to compare with, but, there would be numerous equations, propositions, mathematical proofs and discussions on optimality analysis.

1.1 Brief Problem Statement

This research problem pertains to the motions of the pseudo Euclidean plane - a plane with constant negative curvature. Such a plane is also called hyerbolic plane because the polar representation of a point $\mathbf{a}(a_1, a_2)$ on a pseudo Euclidean plane is given as $a_1 = r\cosh\varphi, \quad a_2 = r\sinh\varphi$, where r is the length and φ is the orientation angle of the position vector of $\mathbf{a}$. Physically such a plane looks like a saddle as given in the figure below:

The motions that we investigate are linear maps of points of the pseudo Euclidean plane that preserve distance and orientation of the points being transformed. Intuitively the motions may be visualized in terms of a unicycle moving on a pseudo Euclidean plane. The unicycle has two degrees of freedom i.e., translation along a line in the plane and rotation about a vertical axis passing through its center of mass. Configuration and state manifold M of the system is 3-dimensional with every point $q = (x, y, z) \in M$ where x, y are two position variables z is the angular orientation variable of the unicycle on the hyperbolic plane. The configuration manifold M alongwith the motions of the pseudo Euclidean plane form a Lie group called special hyperbolic group SH(2).

It is well known that a unicycle is a nonholonomic system which is constrained to have a translational velocity along a line and an angular velocity about its vertical axis. Mathemat-

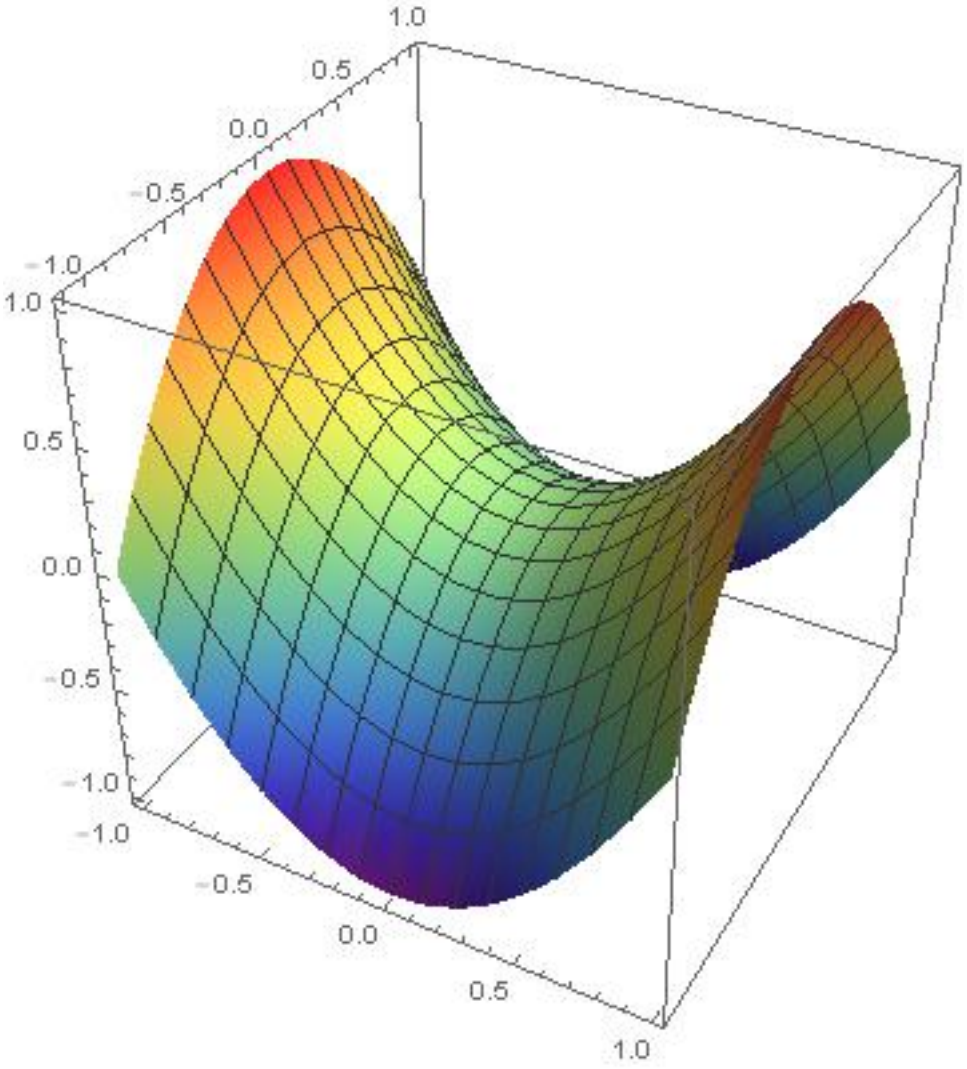

Figure 1.1: Pseudo Euclidean Plane

ically, a unicycle moving on a hyperbolic plane is represented by a driftless control system as:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} u_1 \cosh z \\ u_1 \sinh z \\ u_2 \end{pmatrix}.$$

where u_1 is the translational velocity and u_2 is the angular velocity. Optimal control of a nonholonomic system is mathematically called a sub-Riemannian problem. Hence, the research problem under consideration is the optimal control and optimal trajectory design of a hypothetical unicycle moving on a hyperbolic plane. Given the initial and final states, the objective is to calculate a horizontal curve $\gamma \subset \mathrm{SH}(2)$ between the q_0 and q_1 such that an appropriate cost function $l = \int_0^{t_1} \sqrt{u_1^2 + u_2^2}\, dt \to \min$, called the sub-Riemannian length functional is minimized.

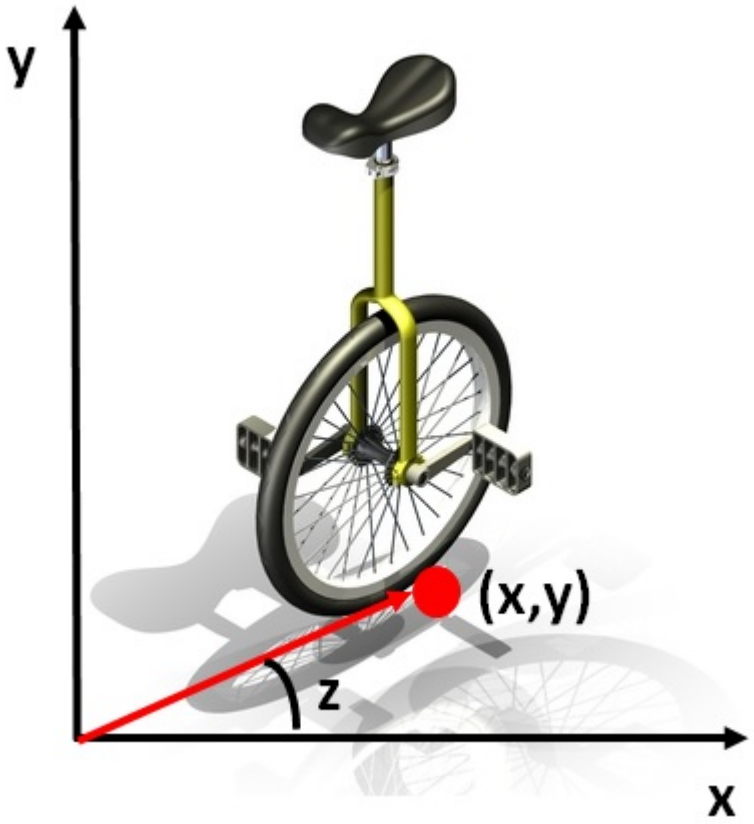

Figure 1.2: Configuration Manifold of a Unicycle

1.2 Motivation

Optimality seems to lie at the heart of all the physical phenomenon in the universe. Amongst trivial examples include catenary, the famous brachistochrone problem, light passing through a medium, image inpainting, the interaction between particles and force fields on quantum scale etc. Human race seems to have developed enormous respect for this principle as every man made process seems to optimize certain cost. Processes that are not optimal in nature evolve and mature to more optimal ones or get weeded out if the optimality is violated. When it comes to optimality in control theory, PMP stands as single most outstanding achievement of mathematics. Discovered purely for an engineering application, PMP is a complete mathematical theory in itself with host of applications that are increasing everyday. It is therefore important and academically rewarding to consider the optimal control of a physical system like unicycle on hyperbolic plane using PMP. There are several other important reasons to consider this research problem i.e.,

- ▷ In the field of robotics, motion is always considered on a Euclidean plane. However, it turns out that the real world space is seldom Euclidean. The space in which a real world object moves always has some curvature, either positive or negative. It is therefore more interesting and relevant to consider the optimal control of such physical

systems on hyperbolic or spherical planes.

- ▷ The technique employed to solve the research problem is geometric control theory. Geometric control is one of the latest advances in control theory whose character is essentially geometric. It is an extension of classical calculus of variations and employs tools of differential geometry, Lie theory etc. to solve the system of differential equations representing the system. This research will utilize, extend and improve the existing techniques available to solve the sub-Riemannian problem via geometric control methods.
- ▷ As will be highlighted later in the thesis, several physical systems such as steering of a car, parking of a car, planar robots and even UAVs can be modeled via equations of a unicycle. Hence, the theoretical results on computation of optimal control and analysis of optimal trajectories is applicable to all such systems.
- ▷ In [2], the authors provided a classification of all sub-Riemannian structures on three dimensional Lie groups in terms of the basic differential invariants. Hence, sub-Riemannian problem on group SH(2) is very important in the entire study of three dimensional Lie groups.

1.3 Research Objectives

The goal of this work is to study the sub-Riemannian problem and address following research problems:

1. Controllability, integrability and existence of optimal trajectories of the dynamical system, [3], [4].
2. Obtain a complete parametrization of extremal trajectories [3].
3. Description of symmetries and the corresponding Maxwell sets [3] [5].
4. Characterize conjugate loci, sub-Riemannian spheres [5].
5. Description of cut loci and global optimality analysis.
6. Description of the global structure of the exponential mapping and the optimal synthesis.

It is pertinent to note that results on research problems 1–4 have been presented in this thesis. The research problems 5 and 6 fall outside the scope of this thesis and have been suggested as future work. As such this is a novel research problem as sub-Riemannian problem on group of motions of pseudo Euclidean plane has not been considered earlier. Optimal control has been computed and local and global optimality analysis has been carried out. A novel orthogonal

transformation has been presented which transforms the sub-Riemannian problem on Lie group SH(2) to an equivalent problem on SOLV^- .

1.4 Thesis Structure

The thesis comprises ten chapters and four appendices including the introduction. Chapter 2, 3 and 4 are our spade work. The whole idea behind these chapters is to provide a short account of essential notions and concepts to the reader and to make the document self sufficient. Hence, brevity has been compromised in favor of completeness. In chapter 2 we collect definitions that are encountered frequently in the thesis and are essential for understanding this work. Definitions are supported by short intuitive explanations and figures wherever possible. Geometric control is an interdisciplinary subject that lies at crossroads of optimal control theory, mechanics, Lie theory etc. In chapter 3 we try to develop the essential connection between these subjects and clarify how geometry, mechanics and control come together. In chapter 4 we briefly explain the concept of sub-Riemannian manifold and then present a detailed literature review. Literature review is multifaceted. It covers equivalent sub-Riemannian problems on various Lie groups, some practical applications of geometric control theory and the numerical integration techniques specially developed for integrating dynamical systems evolving on a Lie group.

Chapter 5 – 9 contain the main results and contributions of this research work. In Chapter 5 we present the detailed problem statement. We also state and explain the objectives of this research. In Chapter 6 we tackle the issues like controllability, computation of open loop control, local representation of the system in terms of canonical coordinates and integrability. We also present an orthogonal transformation that transforms the Lie group SH(2) into the Lie group SOLV^-. In Chapter 7 we compute the Hamiltonian flow and parametrize the extremal trajectories in terms of Jacobi elliptic functions. This parametrization is extremely important for second order optimality analysis that is the subject of Chapters 8 and 9. The second order optimality analysis is basically an elimination process in which we eliminate the trajectories that contain either the Maxwell points or the conjugate points. In Chapter 8 we provide complete description of Maxwell strata based on the symmetries of the vertical subsystem. We obtain the first Maxwell time corresponding to trajectories of the vertical subsystem. The first Maxwell time gives the upper bound on the cut time i.e., time of loss of global optimality. In Chapter 9 we compute the conjugate loci of the extremal trajectories. We compute the Jacobian of the exponential mapping and calculate the first roots of the Jacobian which gives the first conjugate time. We also obtain upper and lower bounds on n-th conjugate times that are important in the description of the sub-Riemannian wavefront.

We conclude the text in Chapter 10 where we describe the objectives achieved in the research work and outline the possibility of future research work in the sub-Riemannian problem on the Lie group SH(2). The Appendix 1 is a detailed description and derivation

of the Jacobi elliptic functions. Appendices 2–4 contain the Mathematica code that was used to compute the Maxwell points, conjugate points and 3-dimensional plots of the sub-Riemannian wavefront and sub-Riemannian sphere.

I feel that in entirety, the subject is extremely beautiful with great potential for further research . Hence, in this thesis I share not my knowledge or my work, but the enthusiasm and the passion that I developed along the way.

Chapter 2

Preliminaries

This research is based on geometric control theory that borrows its concepts from various disciplines of mathematics such as differential geometry, Lie theory etc. Consequently, the research thesis discusses a wide range of topics and concepts that are usually forte of a mathematician and not an engineer. Hence, it is prudent to develop the necessary background and briefly discuss the key concepts that are used later in this work. This chapter is devoted to the definitions and explanations of important terms and concepts discussed later in the text. As such this chapter acts as a ready reference and crystallizes the terminology used in the thesis.

2.1 Configuration Space

The configuration of a dynamical system is its state or position at any given time. The configuration is not related to any motions that the system is executing. So as to say, configuration of a dynamical system is like a snapshot at any time instant which changes when the system moves. Hence the motion of a system is actually series of snapshots taken at infinitesimal time intervals and being seen continuously just like a video that practically comprises of still images called frames. When the frames are changed at certain speed, there is an illusion of moving picture created out of still images. In practice configuration is specified by variables representing the position coordinates of suitably chosen parts of the system, [6].

Definition 2.1.1. The configuration space of the system is therefore an abstract space, whose coordinate points correspond to all physically possible configurations of the system. The configuration space is a manifold whose dimensions are equal to the degrees of freedom of the system [7].

Consider for example a robotic manipulator in Figure 2.1. Its configuration is given by angular position of its revolute and displacement of its prismatic joints. Hence the configuration space is 3 dimensional with $\theta_1, \theta_2 \in \mathbb{R}/2\pi$ and $d_3 \in \{0, l\}$.

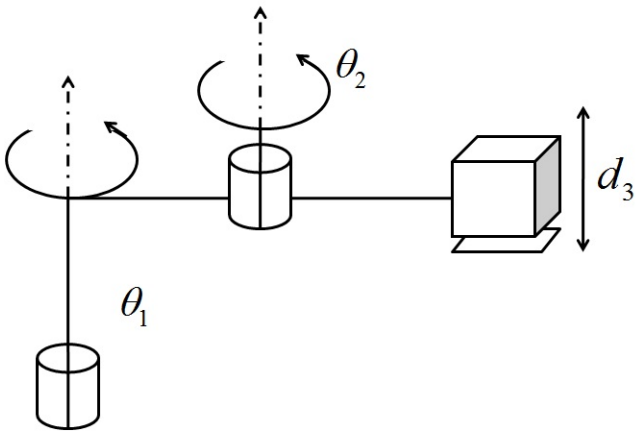

Figure 2.1: Robotic manipulator with 2 revolute and one prismatic joints

2.2 Phase Space

> Phase space has been called one of the most powerful inventions of modern science. But its historical origins are clouded in a tangle of independent discovery and misattributions that persist today.
>
> David D. Nolte

> Hamiltonian Mechanics is geometry in phase space.
>
> Vladimir I. Arnold (1978)

Definition 2.2.1. Phase space is a hypothetical space whose coordinate axis represent the position and velocity states of the system and each point in space corresponds to a particular value of states [8].

Configuration of a mechanical system is an instantaneous description of its geometry. It does not take into account its dynamic response. In contrast, the state of the system is a set of variables that, along with its dynamics and input, completely determines evolution of the system in future. The phase or state space is the set of all possible states. To every n-dimensional configuration manifold there is associated a $2n$-dimensional phase space. First n-dimensions represent the position variables and the other n-dimensions represent the momentum/velocity variables [7]. For example the phase space of the robotic arm presented in Figure 2.1 consists of 3 state variables to represent the position and 3 state variables to represent the velocity of the three links. The velocity variables are the time derivatives of position variables.

For the interested reader, the term phase space evolved over period of more than 5 decades starting from Liouville in 1838. Major role was played by Carl Gustav Jacobi who related Liouville's work on integral equations to dynamical systems and Boltzmann who was

studying the gas phases. Other people who made their mark in development of concept of phase space include Henri Poincare. Finally the term phase space was used by Boltzmann's student Paul Ehrenfest who was reviewing his work in 1911 [8].

2.3 Manifold

Definition 2.3.1. A manifold is a topological space that locally looks like a Euclidean space or that can be locally approximated as Euclidean space but on larger scale it might be much different.

Manifold comes from a Germanic word "*Mannigfaltigkeit*" that simply means "many folds." It was coined by Bernhard Riemann. Literally it means many folds, forms, or features. The wide usage of the word in many disparate fields such as mathematics or exhaust manifolds of vehicles stems from the literal meaning. On cars the exhaust manifolds are structures that have multiple inlets/outlets (hence, many features). Similarly, in topology, manifolds are, in general, surfaces with many features [7],[9],[10].

Manifold Dimensions

The minimum number of independent variables needed to define every point on a manifold is called its dimension. In that respect, the usual geometric objects with n+1 dimensions in Euclidean geometry appear to be n-dimensional manifolds. For example, a circle in Euclidean space is two dimensional figure. From topological point of view every point on the perimeter of a circle is represented by a single variable i.e., the angle θ between horizontal and the radial line at any point. Hence a circle is called 1-manifold. Consequently, there are two ways to visualize a manifold i.e., extrinsic and intrinsic. Extrinsic view is generated when observing the manifold from outside of it as embedded in Euclidean space. Thus n-dimensional manifold embedded in $\mathbb{R}^{n+1}$ is called a 'codimension 1 space'. The main advantage of extrinsic view is the ability to use well developed tools from Euclidean geometry. The intrinsic view of an n-dimensional manifold M on the other hand is an abstract way of considering it as a topological space by itself, without the need to attach notion of Euclidean space with it in order to study its properties. As such this view is generated when the observer dwells on the manifold [7].

Properties

Manifolds may or may not be connected (all in 'one piece'), closed or finite. For example, a pair of separate circles is also a topological manifold (not connected), a line segment without its ends is a manifold (not closed) and a parabola is a topological manifold (not finite) [7].

Types of Manifolds

- ▷ Differentiable or, smooth manifolds, are the ones on which one can do calculus.
- ▷ Riemannian manifolds are manifolds endowed with an intrinsic metric that gives natural measure of distances and by which angles can be defined. In mechanics they serve as the configuration spaces for dynamical systems.
- ▷ Symplectic manifolds endowed with a non degenerate bilinear form and which serve as the phase spaces in mechanics and physics;
- ▷ Pseudo–Riemannian manifolds which model space–time in general relativity.

2.4 Vector Field

Definition 2.4.1. In vector calculus , vector field refers to attaching a vector to every point in the space.

In differential geometry this has more deeper meaning that shall be explained shortly in context of tangent space. Nonetheless, both concepts coincide. Consider for example a pendulum described by differential equations:

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= -\sin(x). \end{aligned} \tag{2.4.1}$$

The phase portrait of the system can be generated by using the well known vector field method. What we essentially do is that at every point in q plane we calculate the value of $\dot{x}$ and $\dot{y}$ and keeping q as the base point we attach an arrow to it with coordinates $\dot{x}$ and $\dot{y}$. When we repeat the procedure for the entire phase plane, we get a bunch of arrows attached to every point in the plane or essentially a field of vectors. The vector field for (2.4.1) is shown in Figure 2.2.

2.5 Flow

The concept of flow is the same as its literal meaning. Consider for example a fluid moving through a volume in space. The motion is called flow. Every particle in the fluid has a velocity with magnitude and direction. If we place a small particle in the fluid it will follow the path and move along the flow. Essentially the particle moves tangentially to a curve in the flow. The tangent to curve is the vector field and hence gives the differential equation of the system. The solution curve is the integral curve as it is obtained by integration of the differential equation or the vector field. The general pattern of arrows in Figure 2.2 gives the path along which a particle will move if placed in it and is called the flow.

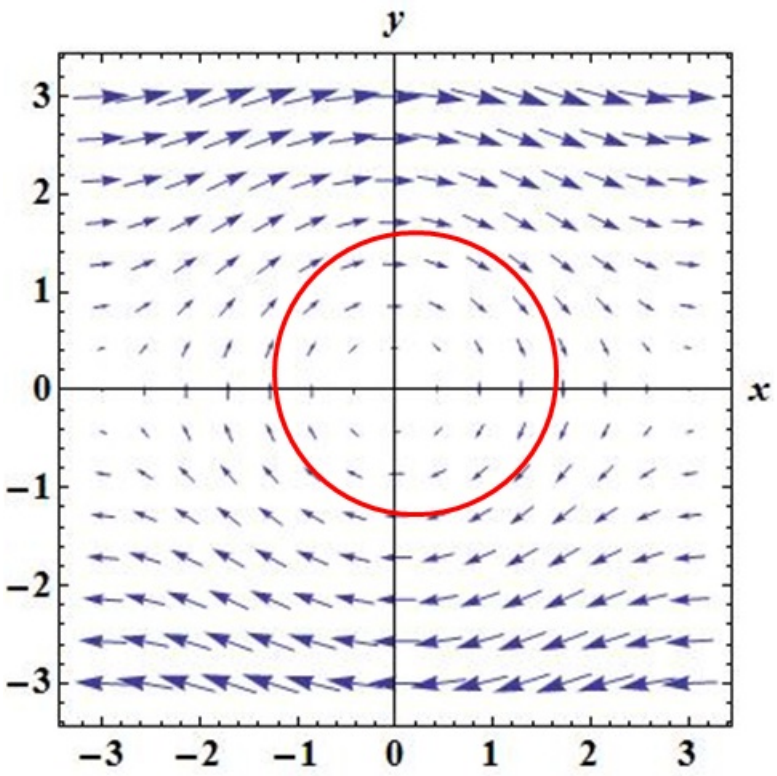

Figure 2.2: Vector Field

2.6 Integral Curve

Definition 2.6.1. A continuous curve $x(t)$ in a manifold M, defined on an interval $[0, T]$, is called an integral curve of family of vector fields F if there exist a partition $0 = t_0 < t_1 < \ldots < t_m = T$ and vector fields $X_1, \ldots, X_m \in F$ such that the restriction of $x(t)$ to each open interval (t_{i-1}, t_i) is differentiable, and $\frac{dx(t)}{dt} = X_i(x(t))$ for $i = 1, \ldots, m$ [11].

The complex mathematical definition amounts to saying only that $x(t)$ is the solution curve of the differential equation given by time-varying vector field $F(x, u(t))$. It is assumed that $u(t)$ is piecewise-constant control taking constant value U_i in each interval $[t_{i-1}, t_i]$. Due to piecewise constant nature of $u(t)$, $x(t)$ comprises of broken continuous curve each piece corresponding to different choices of control values. Note that piecewise constant nature of control input $u(t)$ allows us to apply different control inputs over different intervals and yet being able to control a system that is continuous in nature. Restricting control to being continuous or smooth would render almost every dynamical system uncontrollable in real time. The red curve in Figure 2.2 is the solution curve starting from some point on the curve and is hence called the integral curve.

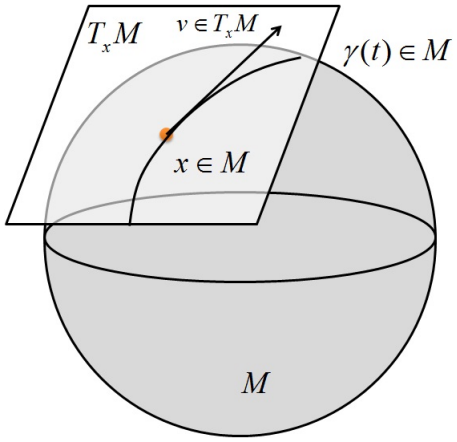

Figure 2.3: The tangent space T_xM and a tangent vector $v \in T_xM$, along a curve $\gamma(t)$ traveling through $x \in M$

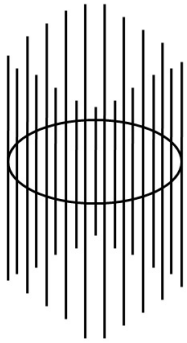

Figure 2.4: The tangent bundle TM - Vertical lines are tangent to 1-sphere at every point and non intersecting

2.7 Tangent Bundle

Definition 2.7.1. Tangent space T_xM to a point x on manifold M is the space containing all curves that pass tangentially from point $x \in M$ [10],[12].

Definition 2.7.2. The disjoint union of all the tangent spaces $T_xM \quad \forall x \in M$ is called the tangent bundle TM.

Thus,

$$TM = \bigsqcup_{x \in M} T_xM. \tag{2.7.1}$$

Tangent space is called fibre of the tangent bundle [10],[12].

Definition 2.7.3. A vector field $f(x)$ is a mapping from manifold M to the tangent bundle TM i.e., $f : M \to TM$ [10].

Definition 2.7.4. The inverse mapping from tangent bundle TM to the manifold M is called projection map $\xi : TM \to M$ [10].

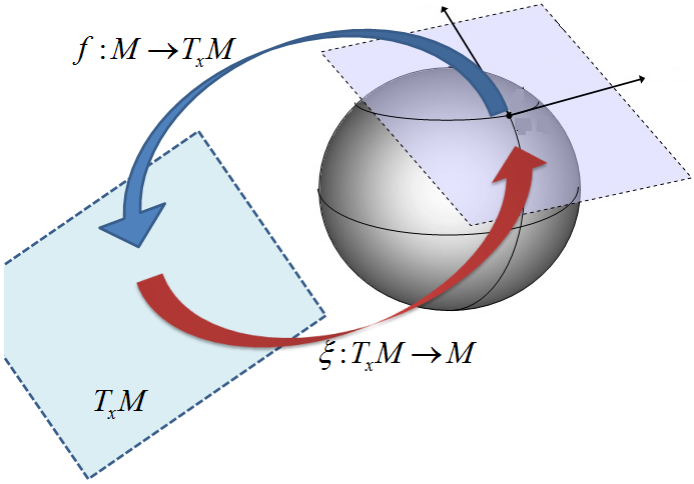

Figure 2.5: The vector field f and projection mapping

The tangent space T_xM at a point $x \in M$ is given as $T_xM = \xi^{-1}(x)$. Note that the RHS of the ODEs that model differential equations are vector fields and the solutions to the ODEs obtained after integration is called projection mapping or exponential mapping. The Lagrangian is a natural energy function on the tangent bundle and tangent bundle is the natural framework for Lagrangian mechanics [7]. Tangent space and tangent bundle are depicted in Figure 2.3 and Figure 2.4 respectively whereas Figure 2.5 represents the idea of vector field and canonical projection.

2.8 Cotangent Bundle

Definition 2.8.1. A covector or 1-form is a linear map from vector space to its field of scalars. For example, the covectors to the column vectors are row vectors [10],[12].

Definition 2.8.2. Cotangent space denoted as T_x^*M is the dual space to tangent space. It is the space of covectors to the tangent vectors [10],[12].

Definition 2.8.3. The union of all disjoint cotangent spaces $T_x^*M \quad \forall x \in M$ is called cotangent bundle T^*M.

$$T^*M = \bigsqcup_{x \in M} T_x^*M. \tag{2.8.1}$$

If the manifold M is n-dimensional the cotangent bundle is $2n$-dimensional i.e., there are $2n$ coordinate axis. First n-coordinates are the position coordinates and other n-coordinates are the momentum coordinates that are the state variables in Hamiltonian mechanics. Hamiltonian is the total energy of the system and is natural energy function on cotangent bundle. Cotangent bundle provides natural phase space for Hamiltonian mechanics. The projection

map of tangent bundle has a dual called conatural projection ξ^* on the cotangent bundle. The cotangent space is the fibre of cotangent bundle [7].

2.9 Group

> *"What is a group? Algebraists teach that this is supposedly a set with two operations that satisfy a load of easily-forgettable axioms. This definition provokes a natural protest: why would any sensible person need such pairs of operations? Oh, curse this maths" concludes the student (who, possibly, becomes the Minister for Science in the future). We get a totally different situation if we start not with the group but with the concept of a transformation (a one-to-one mapping of a set onto itself) as was done historically. A collection of transformations of a set is called a group if along with any two transformations it contains the result of their consecutive application, and along with any transformation its inverse. This is the entire definition. The so-called axioms are in fact just (obvious) properties of groups of transformations* [13]".
>
> V. I. Arnold - On Teaching Mathematics

Definition 2.9.1. An arbitrary set (called carrier set or underlying set) is called algebraic structure if it has one or more finitary operations defined on it [14].

Common examples of algebraic structures include groups, rings, fields, and lattices.

Definition 2.9.2. A group G is an algebraic structure which is closed under a binary operation $\star$ such that for any $g_1, g_2 \in G$, $g_1 \star g_2 \to g_3 \in G$ [7],[15]. It satisfies following properties:

- ▷ Identity - There exists an element $e \in G$ such that $\forall g \in G$, $g \star e = e \star g = g$.
- ▷ Inversion - For every $g \in G$, there exists an inverse $h \in G \quad | \quad g \star h = h \star g = e$.
- ▷ Closure - For all $g_1, g_2 \in G$, $g_1 \star g_2$ is also in G.
- ▷ Associativity - For all $g_1, g_2, g_3 \in G$, $g_1 \star (g_2 \star g_3) = (g_1 \star g_2) \star g_3$.

2.10 Lie Group

Definition 2.10.1. A group G is called Lie group if it is also a smooth manifold and if the linear operation and its inversion operation are smooth as maps of manifolds [15],[7].

The most widely cited example is that of general linear group $GL_n(\mathbb{R})$ which is a group of invertible matrices with matrix product as the group operation. The groups of matrices are also called matrix Lie groups. Lie groups are ubiquitous in mechanics and control and serve as the state manifold of numerous practical systems e.g.,

- ▷ Special Orthogonal Group SO(3) – The group of all rotational motions of a rigid body. Represented by 3x3 orthogonal matrix with determinant 1.
- ▷ Special Euclidean Group on 3D Euclidean space $\mathrm{SE}(3) = \mathrm{SO}(3) \times \mathbb{R}^3$ – The group of 6 DOF rotational and translational motions of a rigid body moving in space. Represented by 4×4 homogenous matrices.
- ▷ Special Euclidean Group on 2D Euclidean plane $\mathrm{SE}(2) = \mathrm{SO}(2) \times \mathbb{R}^2$ – The group of 3 DOF rotational and translational motions of a rigid body moving on a plane. Represented by 3×3 homogeneous matrices.

2.11 Lie Algebra

Algebra is commonly known in the context of branch of mathematics that deals with doing symbolic mathematical computations. However, it has several meanings including the one presented below.

Definition 2.11.1. Field K is an algebraic structure closed under two binary operations of addition and multiplication. Example is that of field $\mathbb{R}$ of real numbers [14].

Definition 2.11.2. Algebra/algebras over a field K is a vector space endowed with a bilinear product $[.,.] : K \times K \to K$. For example algebras over $\mathbb{R}$ is a vector space endowed with bilinear scalar product or the so called dot product [14].

Definition 2.11.3. Lie Algebra $\mathcal{L}$ is an Algebra of Lie Group whose bilinear product $[.,.]$ is also:

1. skew symmetric i.e., $[u, v] = -[v, u] \implies [u, u] = 0$
2. satisfies Jacobi Identity: $[u, [v, w]] + [w, [u, v]] + [v, [w, u]] = 0 \quad \forall u, v, w \in \mathcal{L}$ (Order of operation is important)

The product operation is called the Lie bracket. Other properties of Lie algebra are:

- ▷ The tangent space to a Lie group G at the identity element is called the Lie algebra $\mathcal{L}$ of the Lie group $\mathcal{L} = T_{Id}G$.
- ▷ Lie algebra completely captures the local structure of Lie group

One of the remarkable achievements of Sophus Lie was that he proved that Lie algebra locally represented the Lie group. This fact makes the well developed tools of linear algebra available to Lie theory. In fact the entire concept of manifolds rests on local representation through Euclidean space and existence of smooth transition maps between these local representations called charts [7],[12].

2.12 Commutator

Commutator is known to electrical engineers as a device that is used to convert alternating current into direct current by switching the contacts when the direction of current is getting reversed. The mathematics perspective is remarkably closer. In group theory commutator of a group measures the extent to which the elements of a group commute i.e., change the order or arrangement. Let's clarify the notion with the mathematical definition. The commutator of $g, h \in G$ is given as:

$$[g, h] = g^{-1}h^{-1}gh \tag{2.12.1}$$

What this essentially means is that we perform a group action h, then act on it by g and then do the inverse operations h^{-1} and g^{-1}. If the result is the group identity then the group actions are commutative i.e., going along one first and then second and then coming back in reverse, one reaches the starting point. If a non-identity element of G is produced by the commutator, then the group is not commutative [16].

2.13 Lie Bracket

Lie bracket (the product operation of Lie algebra) is the commutator of vector fields. It is measure of the degree to which vector fields commute. Mathematically it is given as [15]:

$$[f, g](x) = \frac{\partial g}{\partial x} f(x) - \frac{\partial f}{\partial x} g(x), \tag{2.13.1}$$

where f and g are vector fields.

Lie Bracket - Commutator of Vector Fields

The concept and mathematical proof related to Lie bracket as commutator of vector fields needs further elaboration for inquisitive reader and for sake of completeness. Besides being a measure of degree of commutativity of vector fields, concept of Lie bracket is central to defining the nonholonomic systems and establishing controllability of a control system [15],[16],[17]. In this regard we will consider the standard procedure of Taylor expansion of flow of vector fields outlined in [16],[17]. Consider the following control system:

$$\dot{x} = f(x)u_1 + g(x)u_2, \tag{2.13.2}$$

where f and g are vector fields and u_1, u_2 are control variables. We apply the following piecewise constant control input to the system over the interval $[0, 4T]$:

$$u(t) = \begin{cases} (1,0) & \text{for } t \in [0,T), \\ (0,1) & \text{for } t \in [T,2T), \\ (-1,0) & \text{for } t \in [2T,3T), \\ (0,-1) & \text{for } t \in [3T,4T]. \end{cases} \tag{2.13.3}$$

Hence the system moves along $f(x)$ for $t \in [0,T)$, along $g(x)$ for $t \in [T,2T)$, along $-f(x)$ for $t \in [2T,3T)$ and along $-g(x)$ for $t \in [3T,4T]$. The situation is presented in Figure 2.6. The flow of the system for $t \in [0,4T]$ is given as:

$$x(t) = \exp(-Tg)\exp(-Tf)\exp(Tg)\exp(Tf)x_0. \tag{2.13.4}$$

We need to establish the conditions under which the flows commute i.e., starting at $x(0)$ and moving in the manner described above, when does one reach $x(0)$ again. Consider now the Taylor expansion on each time segment T approaching zero so that we can neglect cubic and higher powers of T. For a system $\dot{x} = f(x)$ we use the fact that by chain rule $\ddot{x} = \dot{f}(x) = \frac{\partial f}{\partial x}\frac{dx}{dt} = \frac{\partial f}{\partial x}f(x)$. For $t \in [0,T)$, we have,

$$\begin{aligned} x(T) &= x(0) + T\dot{x}(0) + \frac{1}{2}T^2\ddot{x}(0) + \cdots, \\ &= x(0) + Tf(x(0)) + \frac{1}{2}T^2\frac{\partial f}{\partial x}|_{x(0)}f(x(0)), \end{aligned} \tag{2.13.5}$$

Consider also the Taylor expansion of $g(x(T))$ which is given as:

$$\begin{aligned} g(x(T)) &= g(x(0)) + T\dot{g}(x(0)) + \frac{1}{2}T^2\ddot{g}(x(0)), \\ &= g(x(0)) + T\frac{\partial g}{\partial x}|_{x(0)}\dot{x}(0) + \frac{1}{2}T^2\ddot{g}(x(0)), \\ &= g(x(0)) + T\frac{\partial g}{\partial x}|_{x(0)}f(x(0)) + \frac{1}{2}T^2\ddot{g}(x(0)). \end{aligned} \tag{2.13.6}$$

For $t \in [T,2T)$, we have,

$$\begin{aligned} x(2T) &= x(T) + T\dot{x}(T) + \frac{1}{2}T^2\ddot{x}(T) + \cdots, \\ &= x(T) + Tg(x(T)) + \frac{1}{2}T^2\nabla g(x(T)).g(x(T)). \end{aligned} \tag{2.13.7}$$

Substitute (2.13.5) and (2.13.6) in (2.13.7) and drop evaluation at $x(0)$ from every occurrence of f, g and their derivatives to conserve space.

$$\begin{aligned} x(2T) &= x(0) + Tf + \frac{1}{2}T^2\frac{\partial g}{\partial x}f + T\left(g + T\frac{\partial g}{\partial x}f + \frac{1}{2}T^2\ddot{g}\right) + \frac{1}{2}T^2\frac{\partial g}{\partial x}g, \\ &= x(0) + T(f+g) + T^2\left(\frac{1}{2}\frac{\partial f}{\partial x}f + \frac{\partial g}{\partial x}f + \frac{1}{2}\frac{\partial g}{\partial x}g\right). \end{aligned} \tag{2.13.8}$$

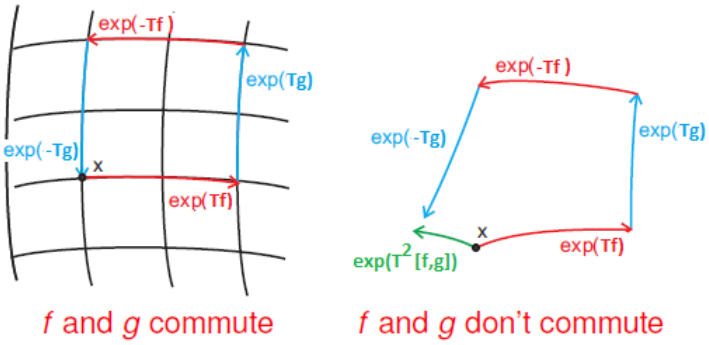

Figure 2.6: Lie Bracket Concept

We repeat the same procedure to obtain $x(3T)$ and $x(4T)$ which is:

$$\begin{aligned} x(3T) &= x(0) + Tg + T^2\left(\frac{\partial g}{\partial x}f - \frac{\partial f}{\partial x}g + \frac{1}{2}\frac{\partial g}{\partial x}g\right), \\ x(4T) &= x(0) + T^2\left(\frac{\partial g}{\partial x}f - \frac{\partial f}{\partial x}g\right). \end{aligned} \tag{2.13.9}$$

From (2.13.9) we see that the state trajectory at $x(4T)$ reaches the initial state or the state identity iff $\left(\frac{\partial g}{\partial x}f - \frac{\partial f}{\partial x}g\right) = 0$ which is by definition the Lie bracket of vector fields f and g. If the Lie bracket $[f, g] \neq 0$, then after the commutator motion shown in Figure 2.6, resulting motion is in the direction which is orthogonal to f and g given as $T^2\left(\frac{\partial g}{\partial x}f - \frac{\partial f}{\partial x}g\right)$. For a distribution to be nonholonomic the Lie commutator has to be nonzero i.e., it has to produce velocities not allowed by the system distribution. If on the other hand $[f, g] = 0$ the vector fields are commutative i.e., if we are trapped on a surface, then it is impossible to leave the surface by using commutator motions. Hence if Lie bracket is not zero, we have a vector field pointing in a direction of motion not given by the original vector fields as it is linearly independent from other set of vector fields. Thus we have an alternative direction of motion available. This notion is extremely important in determining the controllability of systems. Nonholonomic systems can hence alternatively be defined as the ones that have some extra vector fields or directions of motion given by Lie bracket of original vector fields [18].

2.14 Left Invariant Vector Field

Definition 2.14.1. Consider a Lie group G and the corresponding Lie algebra $\mathcal{L}$. Vector fields of the form $V(X) = XA, \quad X \in G, \quad A \in \mathcal{L}$, are called left-invariant vector fields on the linear Lie group G. The Lie bracket of two left-invariant vector fields $V(X) = XA, \quad W(X) = XB \quad X \in G$ is given as [15]:

$$[V, W](X) = [XA, XB] = X[A, B] = X(AB - BA), X \in G. \tag{2.14.1}$$

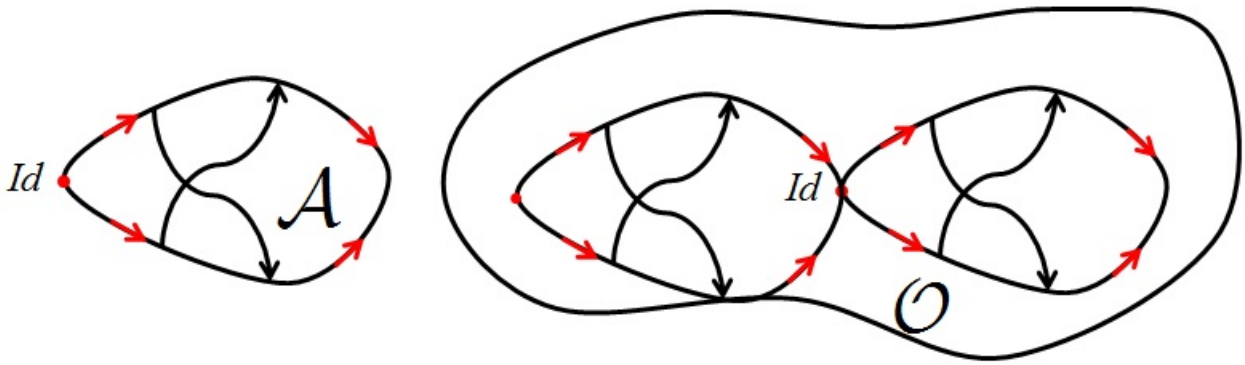

Figure 2.7: Reachable Set

2.15 Left Invariant Control System

Definition 2.15.1. A control system Γ of the form

$$\dot{X} = XA, \quad A \in \mathcal{L}, \quad X \in G \tag{2.15.1}$$

is called left invariant control system [15].

The key property of left invariant control system is that its tangent space T_XG at any point $X \in G$ is the left translated version of the tangent space at the identity obtained by left action of the group on the tangent space i.e., $T_XG = XT_{Id}G$. By virtue of this property the initial conditions for the Cauchy's problem on left invariant control system can be taken as identically zero as any initial conditions $X(0)$ can be left translated to the group identity [15].

2.16 Reachable / Attainable Set

Definition 2.16.1. Consider a left invariant system Γ on Lie group G. Reachable (or attainable) set $\mathcal{A}_\Gamma$ of Γ through a point $q \in G$ consists of all those points in G that can be reached from q in positive time i.e., $t \in \mathbb{R}^+$[15].

2.17 Orbit

Definition 2.17.1. Consider a left invariant system Γ on Lie group G. Orbit $\mathcal{O}_\Gamma$ of Γ through a point $q \in G$ consists of all those points in G that can be reached from q in time t such that $t \in \mathbb{R}$ [15].

Hence $\mathcal{A}_\Gamma \subset \mathcal{O}_\Gamma$ see Figure 2.7. Note that in Figure 2.7 the reachable set is towards one side of the identity of the group G pointing to motion in forward time only, whereas, orbit is towards all sides depicting possibility in negative time. The orbit of family of vector

fields plays an important role in geometric control theory. Geometric control theory takes diversion from rest of the control theory due to Nagano-Sussman orbit theorem [11]. This fact shall be highlighted further in succeeding chapters.

2.18 Geodesic

Definition 2.18.1. A geodesic is a locally length-minimizing curve [19].

In mathematics, particularly differential geometry, a geodesic is a generalization of the notion of a "straight line" to "curved spaces". The term "geodesic" comes from geodesy, the science of measuring the size and shape of Earth. In the original sense, a geodesic was the shortest route between two points on the Earth's surface. In Riemannian geometry, it signifies a curve which is of shortest length at least locally. A curve with shortest length globally is called minimizing geodesic [19].

2.19 Poisson Bracket

Poisson brackets are the Lie bracket analog in the Hamiltonian Classical Mechanics. The Hamiltonian mechanics is defined on the cotangent bundle and the Poisson bracket gives a commutator of the covector fields just like the Lie bracket gives the commutator of vector fields on tangent bundle. Poisson bracket also links the classical and quantum mechanics. Consider system with n particles each with one degree of freedom. The phase space of the system consists of $2n$ coordinates including the n generalized position coordinates $q_1; q_2; ::: q_n$ and n conjugate momenta $p_1; p_2; ::: p_n$. The Poisson bracket is an operation which takes two functions of phase space and time, call them $F(q_i; p_i; t)$ and $G(q_i; p_i; t)$ and produces a new function [20]. It is defined:

$$\{F, G\} = \sum_{i=1}^{n} \frac{\partial F}{\partial q_i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q_i}. \tag{2.19.1}$$

For $n = 1$ this simplifies to:

$$\{F, G\} = \frac{\partial F}{\partial q}\frac{\partial G}{\partial p} - \frac{\partial F}{\partial p}\frac{\partial G}{\partial q}. \tag{2.19.2}$$

One of the reasons for Poisson bracket being a powerful tool is its ability to generate time derivative of a function $F(q, p, t)$ on the cotangent bundle as:

$$\frac{dF}{dt} = \{F, H\} + \frac{\partial F}{\partial t}, \tag{2.19.3}$$

where H is the Hamiltonian function. Intuitively H is the total energy of the system. The time derivatives of the coordinates q, p on cotangent bundle can be simply be obtained by $\dot{q} = \{q, H\} = \frac{\partial H}{\partial p}$ and $\dot{p} = \{p, H\} = -\frac{\partial H}{\partial q}$ which are the equations of motion in Hamiltonian mechanics. Note that Poisson bracket follows the same multiplication rule as the Lie bracket and also satisfies the Jacobi identity. It is also linear in both of its arguments.

2.20 Chapter Summary

This chapter was devoted to building the background that would be necessary in understanding the problem definition and subsequent solution. The concepts related to manifold, differential equation, tangent and cotangent bundle have been covered. However, the material is of preliminary nature and interested reader should look at the references for in depth explanation of the concepts. The next chapters build upon the terminology set forth in this chapter.

Chapter 3

A Primer on Geometric Control Theory

"Mathematics is the language in which God wrote the universe".

Galileo Galilei

It became clear to the mathematicians very early that unlike other disciplines of science, mathematics was universal and was equally valid in any world as in the one we dwell. For example the mathematical constants π or e stay the same whatever universe we may traverse. On the other hand, the universal gas constant R is not as universal as its name sounds. As Stephen Hawking would put it, "indeed if we want to 'read' God's mind, we must first understand God's language". The early mathematician therefore launched the quest to recognize the characters of this universal language. The attempt led to the invention of numbers as a measure of the physical quantities and the discovery of the geometric shapes to describe the physical appearance of the objects. Numbers and geometric shapes have played a fundamental role in the development of human mind and his understanding of the grandeur and beauty of this universe. Numbers gave the meaning to how distant was Marathon from Athens and geometry allowed the geometers to draw straight and parallel lines.

Despite intuitive connection between geometry and numbers, the two mathematical notions and creations of human mind remained disparate and were developed individually. Pythagoras for example studied the rational numbers and were mystified by the structure attached to them [21]. In almost the same era, Euclid was formalizing and laying foundations for Euclidean geometry that would reign the mathematical world for nearly 22 centuries [22]. In 17^{th} century A.D. French philosopher Rene Descartes introduced the Cartesian coordinate system that allowed a point in Euclidean space to be represented as a set of numbers along the coordinate axis. This laid the foundations of Analytical Geometry by which the problems of geometry could be expressed as algebraic expressions and solved easily.

Another unification of knowledge is attributed to Newton who invented and applied calculus to describe the physical phenomenon of motion. Newton not only invented calculus, he also discovered the laws of motion and gravitation [23]. In 19^{th} century, Maxwell formulated

the laws of electromagnetism adding a level of sophistication to mathematics. Einstein's theory of special and general relativity became possible only because the geometry and corresponding mathematics existed to support and explain the theory. The relativity theory as well as electromagnetism were formulated as field theories with underlying four-dimensional space–time, and this fusion of geometry and classical physics provided a strong stimulus to mathematicians in the field of differential geometry.

The famous visions of Rene Descartes in 1619, that led to the invention of algebraic geometry and elucidated the unification of all knowledge through deductive reasoning and mathematics remains the motivation for scientists and mathematicians to-date. Tools from remote branches of mathematics and geometry suddenly become relevant and even essential to the development of new knowledge in other disciplines of science. Nonlinear control theory apparently having much less connection with geometry has witnessed similar transformation. Differential geometry developed in its own right as a subject and not apparently relevant to mechanics and control has become the main tool for the control theorists in recent times. The merger has led to an entire new discipline in control theory known as geometric control theory that allows a more geometric view to the problems of dynamics and control thereby allowing insight that wasn't possible earlier. This work is application of geometric control theory to a novel problem namely the optimal control problem on group of motions of pseudo Euclidean plane. It is therefore imperative to crystallize the connection between geometry, mechanics, control and certain other areas. Needless to emphasize that without clarifying such association, this thesis cannot possibly deliver its outlined objectives.

3.1 Geometry Meets Mechanics

For nearly 22 centuries Euclid's work dominated the mathematical world following its introduction in 3^{rd} century BC. Euclid's axioms symbolize one of the most captivating and emblematic discoveries that laid the foundations of rational science [22]. Euclid's axioms and theorems based on them were intuitively so appealing that it was nearly impossible to visualize and develop a consistent geometry other than Euclidean geometry. Nevertheless, disparate and unsuccessful efforts were made by many geometers to prove the parallel postulate wrong until 19th century. Being the only available notion of space in times of Newton, the state space of a dynamical system being expressed as set of differential equations was naturally the Euclidean space. Thus the notion of state space set out on a wrong track from the onset, though, apparently Euclidean space could successfully explain the physical world and provided a natural setting for the mechanics of 17^{th} century.

In 1788, Joseph Louis Lagrange provided a reformulation of Newtonian mechanics that had its roots in principle of least action and calculus of variation. Lagrange introduced a scalar quantity called Lagrangian that allowed modeling the dynamical equations in generalized coordinates. The Lagrange equations of motion were still the same second order New-

ton's equations, but were far easier to obtain and allowed insight into many new problems. Another reformulation was introduced by William Rowan Hamilton in 1833. Like Lagrangian mechanics, Hamiltonian mechanics also introduced a scalar quantity called Hamiltonian, but the resulting equations were of first order that could be further differentiated to obtain the second order Newton's equations. On the face of it, such reformulation did no better than the earlier formulations. However, it allowed better understanding of the underlying principles. Another bigger advantage was that once the experiments reached on atomic scale, Newtonian and Lagrangian mechanics failed to explain the quantum mechanics. On the contrary, Hamiltonian mechanics could be generalized and extended to quantum mechanics through the introduction of Poisson bracket.

Mechanics was undergoing reformulation apparently unaware of the parallel developments in mathematics and geometry. Efforts to discover non-Euclidean geometries matured when Russian mathematician and geometer, Nikolai Ivanovich Lobachevsky discovered the Hyperbolic geometry [24] and German mathematician, George Friedrich Bernhard Riemann discovered Elliptic geometry [25]. Later in 20^{th} century, Riemann introduced the idea of manifold, a topological space that locally looked like Euclidean space. Impact of discovery of non-Euclidean geometries was profound on physics. It led to the formulation of theory of relativity for which hyperbolic geometry provided the most natural setting. For mechanics, the concept of manifold offered a generic shape of the state space. It was realized that the phase space for Lagrangian mechanics was tangent bundle and the phase space for Hamiltonian mechanics was cotangent bundle [7] that is naturally endowed with non-degenerate symplectic form. What possible difference a non-Euclidean state space can make to our analysis and control design is depicted with example 3.1.1.

Example 3.1.1. Consider a car traveling on the surface of the earth see Figure 3.1. The goal is to reach point P_1 starting from P_0. The configuration manifold for this car is an ellipsoid. Taking usual approach and defining the configuration space as Euclidean space, we fix a coordinate frame at the center of the earth. The position vectors for P_0 and P_1 are $\overrightarrow{v}_0$ and $\overrightarrow{v}_1$ respectively. Vector algebra tells us that in order to reach P_1, the car needs to follow a vector $\overrightarrow{v}_1 - \overrightarrow{v}_0$ traveling a distance of $\|\overrightarrow{v}_1 - \overrightarrow{v}_0\|_2$. Practically, we know that covering the distance of $\|\overrightarrow{v}_1 - \overrightarrow{v}_0\|_2$, the curvature of the earth allows the car to reach only P_2. The error $\|\overrightarrow{v}_1 - \overrightarrow{v}_0\|_2 \to 0$ as $P_0 \to P_1$. What it suggests is that locally any manifold may be approximated as a Euclidean space, however, on macro scale, errors induced by such approximation will blow up. As can be construed, in case of dynamical systems such as aircraft, a ship, a paper rolling mill etc. the ensuing results may be catastrophic. Concept of a non-flat state space revolutionized classical mechanics. With this realization and further development in differential geometry a new term was coined i.e., *"Geometric Mechanics"* which employs differential geometry to describe and study mechanical systems. This exciting new blend of two disciplines is popular and promising area of research with growing number

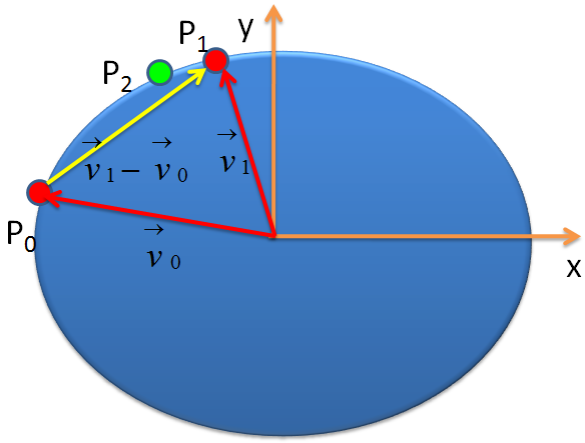

Figure 3.1: Euclidean Framework for Control Theory

of publications every year.

3.2 Geometrization of Control Theory

> "For since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear".
>
> Leonhard Euler

Control theory is relatively young in the entire spectrum of mathematics and optimal control theory is perhaps one of the oldest and most developed parts of it. Optimal control theory sinks directly to the sources of calculus of variations and mechanics and is a subject of interdisciplinary character residing at crossroads for optimal control, differential geometry, and mechanics [26]. Beginning perhaps with the famous brachistochrone problem in 1696 posed by Johann Bernoulli in Acta Eruditorum [27],[28], optimal control theory attained widespread acceptance and attracted plethora of applications following the discovery of Pontryagin's maximum principle (PMP) [26]. Pontryagin and his collaborators were engaged by Soviet air force to design such a trajectory for aircraft that enabled it to get out of a chasing missile's reachable set in minimum time [29]. The calculus of variations was unable to provide answers Pontryagin was looking for. Consequently, Pontryagin formulated his maximum principle laying down conditions for the general time-optimal problem. Discovery of the Maximum Principle was driven by the needs of an optimal control problem and forms classic example where a theory framed to tackle purely engineering problem intractable by existing tools and knowledge, eventually matured into a scientific theory of major significance.

PMP draws its strength from geometry. Like all optimization techniques, PMP provides

conditions for optimality that minimize certain cost. However, there are two discriminating aspects to it. First, the integral cost criterion is minimized along the trajectories of a dynamical system modeled via differential equations. Second, the optimality conditions utilize a fundamental topological fact i.e., the solution to the optimal control problem lies on the boundary of an extended reachable set which is formed by the competing optimal trajectories and their integral costs [20].

Thus PMP provided for the gap that existed between geometry and control [11],[28]. PMP gave birth to an entire new discipline known as geometric control theory that has been subject of extensive research following discovery of PMP. The active research in geometric control provided for the theory that led to extension of the PMP to optimal control problems on arbitrary differentiable manifolds. The theoretical foundation provided by geometric control theory carries important results concerning the differential and topological properties of reachable sets and is an essential component of modern optimal control theory.

3.3 Relevance to Lie Theory

PMP, Nagano-Sussman orbit theorem and a famous theorem from W. L. Chow 1939, set the framework but geometric control started to take shape as an independent discipline of control theory in 1970's following pioneering work from Roger Brockett [26]. Brockett geometrized the nonlinear affine-input control problems by incorporating tools from Lie-theory into his analysis and design. He considered the control system whose state space is a sphere and showed that Lie theory played an important role in analysis of such systems. Following Brockett, there was a bombardment of ideas and mathematics that matured into geometric control theory in 1990s [26].

Imperative here is to understand the connection that Lie theory bears to the control system and what does it offer to a control system designer? Sophus Lie a Norwegian mathematician (1842 - 1899) worked on continuous transformation groups and integration for ODEs via a unified method. His intriguing discovery was to have a far reaching impact on control theory. He established that various integration techniques for ODEs were in fact special cases of a more generalized integration method. This observation was based on the fact that various ODEs were invariant under group action by continuous group of symmetries. Thus solutions of an ODE could be transformed into solution of another ODE once their symmetry group was identified [7]. These groups were later called Lie Groups. As Felix Klein would put it [30]:

> "Geometry is the study of those properties of a set W that remain invariant when its elements are subjected to the transformations of some transformation group Γ".

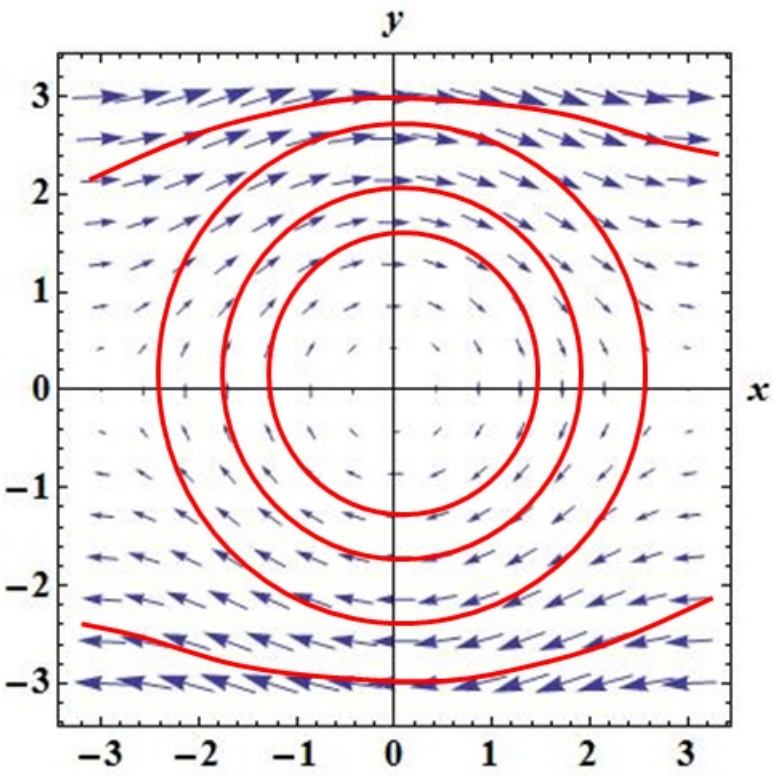

Figure 3.2: Integral Curve and Vector Field

Geometry is thus a double $G(W, \Gamma)$. Transformation groups being at the heart of modern geometry and Lie groups being continuous transformation groups make the connection between them readily apparent. Let us elaborate this concept a bit further using an example. Consider again the equations of pendulum in (2.4.1). The general solution of this set of ODEs is parametrized by constants corresponding to initial states of the pendulum. The family of solution curves along with the tangential vector field are given in Figure 3.2.

The solution curves of differential equations (integral curves) and the corresponding vector fields are essentially geometric objects (they have shape, size and length). The geometric object associated with the differential equations is thus its corresponding field of directions in the plane (the vector field). The field of directions is always tangent to the solution curve corresponding to a particular value of initial conditions. Thus, geometrically, the problem of integration of a differential equation is to find all such curves that are tangent to the given field of directions in the plane. This can be easily seen in Figure 3.2. The solution curves (red) are everywhere tangent to the vector field (arrows).

Symmetry is an important concept of geometry as discussed earlier. Symmetry implies invariance under certain mathematical operation. A differential equation, viewed as a geometric object i.e., field of directions in the plane, may possess some symmetry such as translation etc. The phase portrait of pendulum shows that it has translation and reflection symmetry i.e., the solution curve or phase portrait remains the same even if translated by 2π or reflected about x, y axes or origin. Lie theory tells us that if we know the one parameter symmetry

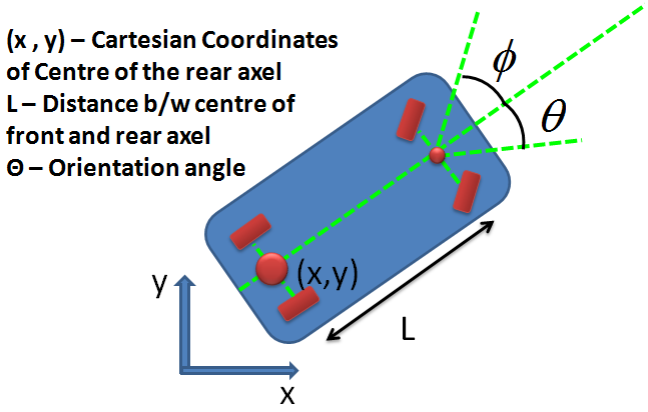

Figure 3.3: Controllability of a Car Steering Model

group of the differential equations we may transform it into a convenient form and solve it. There can be two distinct ways in which Lie theory can help find the solutions of differential equations i.e.,

- ▷ Given a differential equation find its one parameter symmetry group. Alternatively,
- ▷ Given a one parameter symmetry group, find the differential equations that are preserved by this group.

Practically, the first question is more important but also more difficult. Usually answering the second question is more fruitful and easier to find closed form solution of the ODEs. Hence geometry and differential equations are intertwined and closely related. Understanding this linkage is conceptually rewarding and very important in the entire study of geometric control.

Linkage between ODEs and control theory warrants no emphasis. Thus it is not surprising to realize that Lie theory and control are also intimately related. One of Lie theory's major contributions to control theory is the Lie bracket or the commutator of vector fields. Lie bracket generalizes the notion of controllability grammian and thus allows to calculate controllability of control distribution for linear as well as nonlinear systems. In fact the controllability grammian (or Kalman controllability condition) for linear systems or linearized systems sometimes gives erroneous results. Example 3.3.1 is widely reported in literature see e.g. [31] and is given here to explain the claim given above.

Example 3.3.1. The drift free model of steering of car is given as Figure 3.3:

$$\begin{aligned}\dot{x} &= v\cos\theta,\\ \dot{y} &= v\sin\theta,\\ \dot{\theta} &= \omega,\end{aligned}$$

where v is the longitudinal velocity measured at center of real axle (x, y), L is the Lie between the two axles, ω is the steering angular velocity and ϕ is the steering angle. In this

case v and ω are the two control inputs. Let us define the state variables as $x_1 = x$, $x_2 = y$, and input variables as $u_1 = v, \quad u_2 = \omega$. The system in state space is written as:

$$\begin{aligned}\dot{x}_1 &= u_1 cos\theta, \\ \dot{x}_2 &= u_1 sin\theta, \\ \dot{\theta} &= u_2.\end{aligned}$$

If we linearize the system around a point (x_0, y_0, θ_0), we get:

$$\begin{aligned}\dot{x} &= Ax + Bu \\ &= 0x + \begin{bmatrix} \cos\theta_0 & 0 \\ \sin\theta_0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.\end{aligned}$$

Clearly the controllability matrix $M_c = [B, AB] = B$ of the linearized system is rank deficient and therefore the system is not controllable. This contradicts our intuitive understanding that suggests that steering of a car is a controllable problem. Let us apply the notion of Lie bracket. In vector form:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \theta \end{bmatrix} \quad x \in M = \mathbb{R}^2 \times \mathbb{S}^1, \quad g_1(x) = \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix}, \quad g_2(x) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where x represents the state vector, manifold M is the configuration space which consists of position $x, y \in \mathbb{R}^2$and the orientation $\theta \in [0, 2\pi]$. The interval $[0, 2\pi]$ is also known as circle or 1-sphere (sphere of dimension 1) topologically. Note that $g_1(x)$, $g_2(x)$ are the input vector fields and form the control distribution. The overall system is:

$$\dot{x} = u_1 g_1(x) + u_2 g_2(x).$$

Certainly $g_1(x)$ and $g_2(x)$ do not span $\mathbb{R}^2 \times \mathbb{S}^1$. According to Lie theory, the system is completely controllable iff we can find enough good Lie brackets such that the sum of number of linearly independent input vector fields and linearly independent Lie brackets is equal to the dimension of the state space. In the example above we have 2 linearly independent vector fields whereas the state space is 3-dimensional. Hence, with two linearly independent vector fields we can span 2-dimensional space only and we need one more vector field that is linearly independent of the other two to span the entire state space. The additional velocity vector field or the direction of motion is given by the Lie bracket to span the entire state space $\mathbb{R}^2 \times \mathbb{S}^1$. The Lie bracket is given as:

$$[g_1, g_2](x) = -\begin{bmatrix} 0 & 0 & -\sin\theta \\ 0 & 0 & \cos\theta \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{bmatrix}.$$

which is linearly independent of the other two basis vectors of the control distribution. Hence $g_1(x)$, $g_2(x)$ and $[g_1, g_2]$ span $\mathbb{R}^2 \times \mathbb{S}^1$ and hence the system is controllable.

Example 3.3.1 depicts how geometric control theory aided with Lie theory can offer alternative view and deeper insight into the existing problems. Consequently geometric control theory has inspired mathematicians and control practitioners around the globe. Besides the problems it has addressed and the results it has achieved so far, it has taken shape of a theory with strong and definite personality that is already proffering valuable returns to its ancestors.

3.4 The Orbit Theorem

The famous Nagano-Sussman theorem gives geometric control theory its very character that is distinct from rest of the control theory. In simple terms it says that the orbit of a family of vector fields is a connected submanifold of the manifold M. This ensures two things i.e., we are on the same configuration manifold and second the orbit is not a Euclidean space it is the same kind of manifold as M. This means we parted our ways with traditional theory of dynamical systems in which state space is Euclidean.
It is trivial to see that any system is controllable if its orbit is the complete manifold i.e., if we can reach all points on the manifold, the system is controllable which is inline with the definition of controllability taught in elementary linear system theory. And finally by Chow's theorem we say that the orbit is a complete manifold M if the Lie algebra of the family of vector fields spans the entire tangent space. Lie theory gives us the tool i.e., Lie bracket to check whether the family of vector fields spans the entire tangent space. Hence these three facts set all the framework needed to study the dynamical system using tools from differential geometry.

3.5 Geometric Control - An Engineer's Stake

Geometric control theory establishes the link between control theory and geometric mechanics i.e., the link between control theory with a geometric view of classical mechanics in both its Lagrangian and Hamiltonian formulations. Mechanics has traditionally described the behavior of free and interacting particles and bodies, the interaction being described by potential forces. Control theory on the other hand provides tools and techniques for prescribing desired motion and providing control over the system dynamics to achieve the desired motion. Hence, mechanics is about describing and control is about prescribing. Modern control theory began largely as linear theory, having its roots in electrical engineering using linear algebra, complex variable theory, and functional analysis as its principal tools. Traditionally, control of nonlinear systems has relied on linear techniques such as linearization about equilibrium points, feedback linearization where we cancel out the nonlinearities via state feedback, sliding mode control techniques where we force the actual system to attain the dynamics of our choice etc. Since almost all real world systems are practically nonlinear in

nature, therefore the performance of linear techniques for nonlinear system control is rather limited. On the other hand the design technique such as feedback linearization is unable to cater for the model variation arising out of simplifications introduced during the modeling process. Similarly, sliding mode control despite being robust to model variations has its own limitations such as chattering. Hence, there is a vast margin of performance improvement in nonlinear control if some technique which is inherently nonlinear in character is available. Starting in 1970's with the pioneering work of Roger Brockett, geometric control theory can offer such dividends, primarily because it is inherently developed in the language of nonlinear systems. It relies on differential geometry which introduces the notion of non-Euclidean spaces called manifolds and works through the problems in coordinate free language. Hence, from the perspective of analysis and design, geometric control theory is a complete recipe for the optimal control of nonlinear systems. Sub-Riemannian problem on the other hand is particularly suited to optimal control of nonholonomic systems whose cost under consideration is the sub-Riemannian length. With this realization, vast array of tools from sub-Riemannian geometry become available to foray analysis and control design of nonholonomic systems. In that respect, geometric control theory has much to offer to control engineers and may be in few years time it will become essential part of control theory curriculum.

3.6 The Big Picture

In order to gain insight and better comprehension of optimal control in general and PMP in particular, it is perhaps prudent to briefly look at the entire spectrum of optimal control and various notions encountered therein. Despite the acclaimed and acknowledged prominence of optimal control in engineering and science, the mathematical jargon is less comprehensively understood without reading couple of texts on each topic. The drawback of such approach is that the bigger picture and the corresponding position of various tools of optimal control theory remains vague at best. We therefore present a brief overview of notion of optimality and optimal control techniques other than PMP for the interested reader. This exposition is by no means exhaustive or an alternate for a complete volume on each of these topics.
There are two fundamental ideas upon which optimal control theory is established i.e., PMP and Dynamic Programming (DP) with allied optimality principle [32]. While Pontryagin and his compatriots were strengthening foundations of PMP, Richard Bellman from USA was independently developing a parallel approach to optimal control that he later termed as dynamic programming. While PMP is a variational approach, DP is based on principle of optimality (to be discussed later) [32]. Looking at the canvas of optimal control one finds following categorization of optimal control techniques:

1. PMP

    a) Applies to deterministic systems only

b) Special case - Linear quadratic control

c) Optimal estimation is closely related to optimal control. Most widely used optimal estimator is Kalman filter.

d) Optimal estimation and linear quadratic control are dual problems

2. Dynamic programming

a) Applies to deterministic, continuous as well as discrete time and combinatorial systems (systems with quantized states and controls).

b) Discrete case is DP whereas continuous case leads to Hamilton Jacobi Bellman (HJB) equation.

c) Generalizes to stochastic systems (systems in which the certain input u when applied to system at some state q_p always yields the same next state q_n are deterministic while those whose next state q_n is only probabilistically related to the previous state q_p and control input u are stochastic).

Discussion on PMP and its allied mathematics shall be dealt in detail in chapter 5, Theorem 6.4.1. Other notions of optimal control are given in succeeding paragraphs for interested readers.

Linear Quadratic Control

From 1960-1961, Rudolf E. Kalman and his collaborators, working in USA, published three seminal papers. In one of these papers Kalman presented the design equations for linear quadratic (LQ) control. Although discovered independently, LQ control can be derived as a special case of PMP when cost functional is quadratic and the dynamical equations of the system are linear [33]. Consider for example a system in state space form:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(t_0) = x_0, \quad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^m,$$

where A is the system matrix and B is the input matrix. The performance index for LQ control problem is given as:

$$J = \frac{1}{2}x^T(t_1)Sx(t_1) + \frac{1}{2}\int_{t_0}^{t_1} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt,$$

where t_0, t_1 are initial and final time respectively, S and Q are positive semidefinite matrices and R is a positive definite matrix. In case of infinite horizon problem i.e., $(t_1 \to \infty)$, the matrix $S = 0$. Assume that there exists a stabilizing feedback controller K of the form $u = -Kx$ that minimizes the cost functional J. Such a feedback controller exists if the solution P to Algebraic Riccati equation:

$$A^TP + PA + Q - PBR^{-1}B^TP = 0,$$

is positive definite. The controller K in this case is given as:

$$K = R^{-1}B^T P.$$

Such a stabilizing controller is called linear quadratic regulator (LQR).

Optimal Estimation and Kalman Filter

In third of his papers mentioned in the preceding paragraphs, Kalman discussed the concept of optimal filtering and estimation theory and proposed the mathematics and design procedure for the discrete Kalman filter. The continuous counterpart of the discrete Kalman filter was proposed by Kalman and Bucy later in 1961 [34]. Also termed as linear quadratic estimator (LQE), Kalman filter is a recursive algorithm that takes the series of state values as input measured from sensors and possibly smeared in noise, and produces an estimate of the current states. The notion affords major implications in control problems particularly state feedback design procedures where it may not be either possible or feasible to measure all states. It is also referred to as Stratonovich–Kalman–Bucy filter as it turns out to be a special case of a more general, non-linear filter proposed earlier by the Soviet mathematician Ruslan L. Stratonovich.

Kalman filter's strength stems from the fact that it can predict past, present and even future states. The whole algorithm is based on cyclic process with two sub-processes i.e., time update or predict and measurement update or correct. Intuitively, the filter generates an estimate of the state ahead of time in the predict sub-process based on previous estimate and without the knowledge of measurement from sensors. Such an estimate is called *a priori* estimate. In the measurement update sub-process, the filter incorporates the available measurements into the a priori estimate to form a better and corrected estimate of the state known as *a posteriori* estimate. The measurement update therefore acts more like the feedback in the filter. The equations describing the filter are briefly discussed below for the sake of completeness [34].

Consider a linear time-invariant discrete time control system:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_{k-1} + w_{k-1}, \qquad x_k \in \mathbb{R}^n, \\ y_k &= x_k + v_k \qquad y_k \in \mathbb{R}^m, \end{aligned}$$

where w_k and v_k represent process and measurement noise respectively assumed to be white and Gaussian. Let $\bar{x}_k$ be a priori and $\hat{x}_k$ be a posteriori state estimates respectively. The state estimate is given as:

$$\hat{x}_k = \bar{x}_k + K(y_k - C\bar{x}_k).$$

Intuitively, it means that we make a priori estimate $\bar{x}_k$ and then make a posteriori estimate by imparting correction to $\bar{x}_k$ based on difference $(y_k - C\bar{x}_k)$ between the measurement at

instant k and the prediction $C\bar{x}_k$. The $n \times m$ matrix K is the filter gain. In order to calculate K consider the a priori and a posteriori errors respectively as:

$$\begin{aligned} \bar{e}_k &= x_k - \bar{x}_k, \\ \hat{e}_k &= x_k - \hat{x}_k. \end{aligned}$$

Corresponding error covariance matrices are given as:

$$\begin{aligned} \bar{P}_k &= E\left[\bar{e}_k \bar{e}_k^T\right], \\ \hat{P}_k &= E\left[\hat{e}_k \hat{e}_k^T\right]. \end{aligned}$$

The filter gain K is such that it minimizes the a posteriori covariance $\hat{P}_k$. It is obtained from another Riccati equation which is not noted here being out of scope of this document. One form of K that minimizes the covariance is given as [34]:

$$\begin{aligned} K_k &= \bar{P}_k C^T \left(C \bar{P}_k C^T + R\right)^{-1}, \\ &= \frac{\bar{P}_k C^T}{C \bar{P}_k C^T + R}, \end{aligned}$$

where R is the measurement noise covariance matrix. Similar result holds for continuous time systems and is call Kalman Bucy filter. Kalman filter is extremely powerful estimator but it is applicable to linear systems only. Much of Kalman filter's success is attributed to non-linear applications though. In order to achieve that, system is linearized locally about the current state with current mean and covariance. In this case the filter is referred to as Extended Kalman Filter (EKF).

Duality Between Optimal Estimation and Kalman Filter

Together LQR and Kalman filter solve the linear quadratic Gaussian (LQG) problem. However, as described by Kalman, LQR and Kalman filter are dual problems of each other i.e., one can be translated into other [32]. This can be established by looking at the Riccati equations for regulator and controller available in standard text. The equations are not given here as considered beyond the scope of this document.

Dynamic Programming

Dynamic programming was introduced as an alternate optimal control technique by Richard Bellman in 1957 [32],[35]. The word programming originally did not bear any relationship to software programming though it represents an important class of optimization algorithms in software engineering now. Essentially, dynamic describes the evolving nature of systems with time and programming means planning. Thus dynamic programming truly means planning optimal actions over time varying states of the system. Dynamic programming is

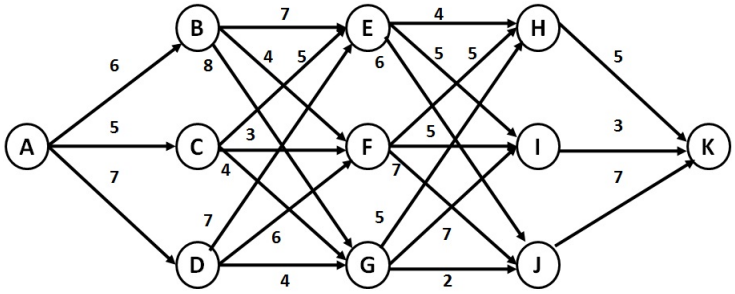

Figure 3.4: Dynamic Programming - Shortest Path

an optimization method for discrete time systems based on Bellman's principle of optimality which in simple terms states that any segment of the optimal trajectory is itself optimal. It highlights the fact that natural direction for solving optimization problems is reverse i.e., starting from the terminal state and working the way backwards through the previous states such that the cost to go in reverse is minimized. As the algorithm reaches the initial state the optimal path with allied cost and optimal control is identified. The cornerstone of dynamic programming is the optimal value function $v(x)$ which is the minimal total cost for achieving the final state x_f starting from the given initial state x_0. We present a simple example motivated from [36] to explain the algorithm.

Example 3.6.1. Consider you wish to travel from a city A to city K representing states x_0 and x_{10} respectively as shown in Figure 3.4. All other towns ordered alphabetically represent states $x_i, \quad i = 1, \ldots, 9$.

The roads connecting A to K pass through various towns. The length of the road in km from town to town is shown in Figure 3.4. The dynamic programming problem is to choose an action u (an appropriate road) at every node x_i that minimizes the total distance traveled (cost) in going from A to K.

1. DP starts at node K and looks for all roads leading directly from other towns to K. In this case these are H, I, J.

2. The algorithm calculates the cost in going from each of these towns to K and attaches this to the nodes as their value $v(x)$ or the optimal cost to go from *i-th* town to K. In this example $v(x_7) = 5$, $v(x_8) = 3$ and $v(x_9) = 7$.

3. Working backwards, we now move on to the previous stage of towns i.e., E, F and G. Considering town E (x_4) first, we calculate its value $v(x_4)$ which is the sum of the shortest running cost in going from x_4 to adjoining town $x_i, \quad i = 7, 8, 9$ in the next stage and the value function of the town x_i. The path that leads to the shortest value is chosen as value of town E ($v(x_4) = 8$ in this case). The situation is explained in Figure 3.5.

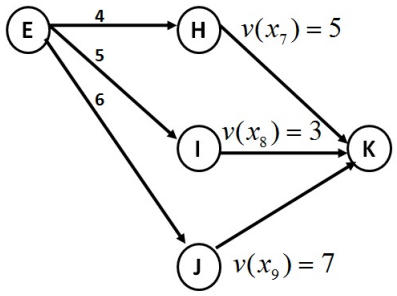

Figure 3.5: Shortest Path from E to K

4. Continuing in this fashion we work backwards ultimately reaching city A (x_0).

5. Suppose that at any state x_i, the value function to go from state x_{i+1} to x_f is $v(x_{i+1}, u)$ and the cost to reach state from x_i to x_{i+1} is $C(x_i, u)$. The optimal control function and corresponding value function are given as:

$$\mathbb{U} = arg \min_{u \in \mathcal{U}(x)} \left(C(x, u) + v(x, u) \right), \tag{3.6.1}$$

$$v(x) = \min_{u \in \mathcal{U}(x)} \left(C(x, u) + v(x, u) \right). \tag{3.6.2}$$

Equations (3.6.1)–(3.6.2) are called Bellman equations.

Hamilton Jacobi Bellman Equations

The principle of optimality applied to continuous time system leads to some very complex partial differential equations known as Hamliton-Jacobi-Bellman equations as they are based on the work done by W. R. Hamilton and Carl Jakob Jacobi. Consider a system whose dynamics and the cost to go are given as:

$$\begin{aligned} \dot{x} &= f(t, x, u), \\ J &= \int_t^{t_f} L(t, x, u)dt + K(x(t_f)), \end{aligned}$$

where t is the initial time and t_f is the terminal time $x(t_f)$, L is the running cost in reaching $x(t_f)$ and K is the possible terminal cost associated with the terminal state x_f. The value function is given as:

$$V(t, x) = \inf_{u \in [t, t_f]} J(t, x, u(.)).$$

Note that value function for initial state $v(t, x_0)$ is the optimal cost in going from x_0 to x_f and is essentially equal to $K(x(t_f))$. Note also the time dependence of value function. The optimal cost functional J^* is given as:

$$J^* = \min_{u \in [t, t_f]} \left\{ \int_t^{t_f} L(t, x, u)dt + K(x(t_f)) \right\}. \tag{3.6.3}$$

Using additive property of integrals we may write equation (3.6.3) as:

$$J^* = \min_{u\in[t,t+\Delta t]} \left\{ \int_t^{t+\Delta t} L(t,x,u)dt + J^*\left(t+\Delta t, x(t+\Delta t)\right) \right\}. \tag{3.6.4}$$

Thus using equation (3.6.4) we have reduced the problem of finding the optimal control over entire interval $[t, t_f]$ to interval $[t, t+\Delta t]$. Note the similarity with the discrete case where we worked out the optimal decision from state to state instead of doing it over entire path from A to K. Equation (3.6.4) may be approximated as:

$$J^*(t,x) = \min_u \left\{ L(t,x)\Delta t + J^*\left(t+\Delta t, x(t+\Delta t)\right) \right\}. \tag{3.6.5}$$

Since $\Delta x = x(t+\Delta t) - x(t) \implies x(t+\Delta t) = x(t) + \Delta x$, the Taylor expansion of $J^*\left(t+\Delta t, x(t+\Delta t)\right)$ ignoring second and higher order terms may be written as:

$$\begin{aligned} J^*\left(t+\Delta t, x(t+\Delta t)\right) &= J^*(t,x) + \left(\frac{\partial}{\partial t}J^*\right)^T \Delta t + \frac{\partial}{\partial x}J^*\Delta x, \\ &= J^*(t,x) + \frac{\partial}{\partial t}J^*\Delta t + \left(\frac{\partial}{\partial x}J^*\right)^T \frac{\Delta x}{\Delta t}\Delta t. \end{aligned} \tag{3.6.6}$$

Substituting equation (3.6.6) into equation (3.6.5) we get:

$$J^*(t,x) = \min_u \left\{ L(t,x)\Delta t + J^*(t,x) + \frac{\partial}{\partial t}J^*\Delta t + \left(\frac{\partial}{\partial x}J^*\right)^T \Delta x \frac{\Delta t}{\Delta t} \right\},$$

$$0 = \min_u \left\{ \left(L(t,x) + \frac{\partial}{\partial t}J^* + \left(\frac{\partial}{\partial x}J^*\right)^T \frac{\Delta x}{\Delta t} \right) \Delta t \right\}. \tag{3.6.7}$$

In (3.6.7) dividing both sides by Δt and then taking $\lim \Delta t \to 0$, we get:

$$0 = \frac{\partial}{\partial t}J^* + \min_u \left\{ L(t,x) + \langle \frac{\partial}{\partial x}J^*, f(t,x,u)\rangle \right\}. \tag{3.6.8}$$

Equation (3.6.8) is the famous Hamilton Jacobi Bellman equation that gives sufficient condition for optimality. Note that it is PDE in time t and state x. The solution to HJB equation gives the optimal control u^*. Note that since value function is the optimal cost J^*, same equation may be written in terms of value function.

3.7 Comparison between PMP and HJB Equations

1. PMP

   a) Gives first order necessary optimality conditions.

   b) The resulting equations are first order ODEs that are relatively simpler to solve.

   c) Candidate optimal trajectories are found.

d) The approach is closely knitted to geometry leading to better insight into the system dynamics.

e) Analytical solution to PMP equations is generally possible.

f) Results in open loop control.

2. HJB Equations

a) Gives sufficient conditions for optimality.

b) Resulting equations are PDEs in time t and state x.

c) The value function has to be known at the start of the procedure which is quite unusual.

d) Results in feedback control.

e) Suffers from curse of dimensionality i.e., the algorithm complexity increases tremendously with increase in the system states n

f) Usually HJB equation has non-smooth solution, and several theories were constructed for generalized solutions to HJB equation (e.g. viscosity solution).

Due to these reasons, applying PMP is usually much easier than applying HJB. Certainly, if one succeeds to solve HJB, this would give a complete solution to an optimal control problem. But this can be done in very rare cases.

3.8 Final Pieces of Big Picture

Model Predictive Control (MPC)

MPC has been around in the process industry since 1980s and has recently experienced a resurgence of interest and new applications. It is also an optimal control technique based on iterative finite horizon optimization of plant model. Number of plant models (usually linear obtained through system identification) are obtained in the entire envelope of operation of the system. At any time t, the state $x(t)$ of the system is sampled and an appropriate model is chosen from a set of available models based on the current state $x(t)$. During the interval $[t, t+T]$, an optimal control is computed by solving Euler-Lagrange equations online that is valid until $t+T$. Once T has elapsed, the system state is sampled again and the procedure is repeated. If there is no model available matching the current state of the system, then, an appropriate interpolation technique is used to provide a suitable model to be used for calculation of optimal control.
As can be seen, the technique is not entirely optimal and it is optimal over finite horizon only. Hence the optimality over infinite horizon is not guaranteed. One could argue that it is

a control scheme that is motivated from engineering applications rather than mathematical theory. Interestingly, finite horizon version of PMP can be applied in MPC settings [32].

Markov Decision Processes

An optimal control problem with discrete states and actions and probabilistic state transitions is called a Markov decision process (MDP). MDPs are extensively studied in reinforcement learning which is a sub-field of machine learning focusing on optimal control problems with discrete state. Dynamic programming easily generalizes to MDPs because of close resemblance in optimality principle and the Markov property of processes i.e., future state is conditionally independent of the past given the present state. The Bellman equations obtained stay the same except that state transitions and value function are now probabilistic rather than deterministic [32].

3.9 Chapter Summary

Geometry, control theory and Lie theory make a recipe that carries potential to bring about a revolution in how we deal with nonlinear systems. Differential equations viewed as geometric objects give a totally different meaning to dynamical systems not possible in the classical theory of integration of differential equations. Alongside geometrization of mechanics and control theory, optimal control theory reached its zenith with the discovery of PMP. The blend of these concepts gave rise to entire new discipline called geometric control theory. PMP and HJB equations are two parallel approaches to optimal control with peculiarities related to each one of them. It turns out that PMP with its geometric view not only solves the optimal control problem, it gives far greater insight into the qualitative behavior of the system.

The notion of optimal control has been firmly established and competing directions have been explored. With motivation for geometric control and PMP constituted, we are now in a position to present the Sub-Riemannian problem which is the main goal of this research.

Chapter 4

Literature Review - Sub-Riemannian Geometry and Geometric Control

In the previous chapter we looked at the chronological development and essential ingredients of the geometric control techniques. We also presented the motivation behind development of the subject and through an example we presented its advantages in establishing controllability of the control distribution. We also presented various optimal control techniques and established why PMP and optimality analysis via geometric control theory is more interesting and rewarding. With the essential mathematical structure in perspective, we now define the sub-Riemannian manifold that forms the configuration manifold of our research problem. We also present a review of the existing literature on the analysis techniques and results on sub-Riemannian problem on various Lie groups.

Definition 4.0.1. Metric Tensor

A metric tensor g_p is a non-degenerate, symmetric, smooth, bilinear map which maps a pair of tangent vectors at each tangent space T_pM of the manifold to a real number [19]:

$$g_p : T_pM \times T_pM \to \mathbb{R}. \tag{4.0.1}$$

4.1 Riemannian Manifold

Carl Friedrich Gauss sometimes referred to as "greatest mathematician since antiquity" proved his Theorema Egregium in 1821, identifying an intrinsic property of curvature of surfaces. Roughly speaking he established that the curvature of a surface completely determined the distances along paths on that surface and is not dependent upon how the surface is embedded in 3-dimensional space. Few decades later, Riemann working on non-flat higher dimensional spaces extended Gauss's idea. He defined a metric tensor that allowed measuring distances and angles on the manifold. The Riemannian tensor defined the notion of curvature

as intrinsic property to the manifold that did not depend upon embedding of n-dimensional manifold in higher-dimensional spaces. The manifolds endowed with Riemannian tensor are called Riemannian manifolds [19].

Mathematically, a Riemannian manifold (M, g) is a real smooth manifold M endowed with an intrinsic metric $g_p \ \forall p \in M$ [37],[38], that can be used to define the notions of length of curves, distance between points and angles on Riemannian manifolds. The metric is non-degenerate and positive definite and is called Riemannian metric or Riemannian tensor [39]. As such Riemannian tensor is a generalization of the concept of inner product on Euclidean space.

4.2 Sub-Riemannian Manifold

Definition 4.2.1. Distribution

A distribution Δ of a manifold M is a sub-bundle of its tangent bundle TM i.e., $\Delta \subset TM$. It is a family of linear subspaces Δ_q of tangent spaces T_qM depending smoothly upon the point $q \in M$ i.e.,$\{\Delta_q \subset T_qM \quad \forall q \in M\}$ [38]. The dimension or rank m of the distribution is equal to the number of independent basis vector fields in the distribution. Sub-Riemannian manifold is a generalization of Riemannian manifold. A sub-Riemannian

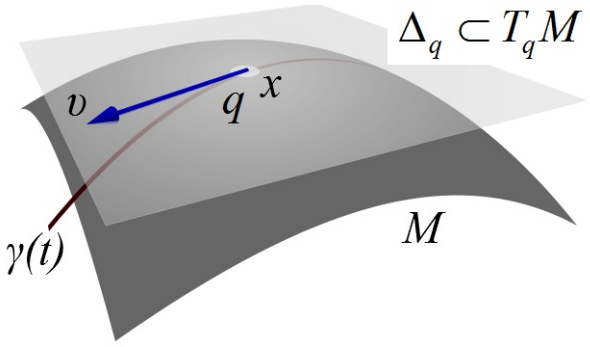

Figure 4.1: Generic Concept of Distribution - The tangent space T_qM at q is 3-dimensional but Δ_q is a plane i.e., it is 2-dimensional

manifold is a triple (M, g, Δ) i.e., n-dimensional Riemannian manifold M endowed with a smooth vector distribution Δ of constant rank $m \mid \quad m < n$ and positive definite Riemannian metric g. The Riemannian metric g is used to define the notion of sub-Riemannian distance d which is a parametrization independent measure of distance between two points q_1 and q_2 on the sub-Riemannian manifold. On a Riemannian manifold the curves have locally finite length but in sub-Riemannian manifold, the curves have infinite length if they do not comply to the constraint on direction of motion. Thus in order to measure distance one is allowed to move along the curves tangent to the so called horizontal subspaces. Figure 4.2 shows a sub-Riemannian manifold M with the distribution Δ. The manifold is three dimensional and therefore its tangent bundle TM comprises non-overlapping three dimensional vector spaces

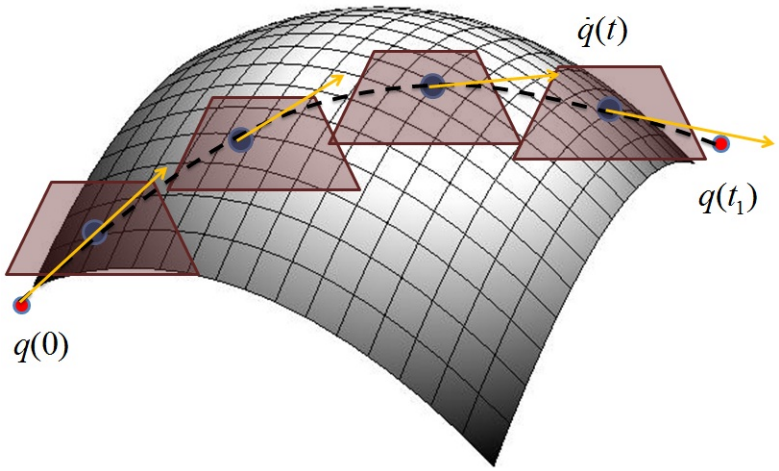

Figure 4.2: Sub-Riemannian Distribution

T_qM at every $q \in M$. The distribution $\Delta \subset TM$ is on the other hand two dimensional which essentially suggests that in order to traverse the manifold M, one is allowed to take velocities in distribution only i.e., $\dot{q}(t) \in \Delta$. The distance in going from $q(0)$ to $q(t_1)$ is given by the smoothly varying positive definite quadratic form g.

Sub-Riemannian Distance

The notion of sub-Riemannian distance is central to the optimal control problems on sub-Riemannian manifolds and needs explanation. Consider a sub-Riemannian manifold (M, Δ, g) and a Lipschitzian horizontal curve $\gamma : I \subset \mathbb{R} \to M; \dot{\gamma}(t) \in \Delta_{\gamma(t)}$ for almost all $t \in I$. The metric g being an inner product in the distribution $\Delta_{\gamma(t)} \subset TM$, the length of curve γ on the manifold M is given as [38]:

$$length(\gamma) = \int_I \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt.$$

Put simply, γ is a curve that maps an interval I to the manifold M such that the $\dot{\gamma}(t)$ is everywhere tangent to the manifold and lies in the distribution $\Delta \subset TM$. If we combine the infinitesimal versions of $\dot{\gamma}(t) \quad \forall t \in I$, we essentially have the whole curve γ. If we calculate the inner product of each individual tangent vector $\dot{\gamma}(t)$ with itself such that $t \to 0$, we have a quantity which is the equivalent to Euclidean norm of vector representing its length. When we integrate lengths of all infinitesimal tangent vectors over the whole interval I, we have the length of the curve γ. The sub-Riemannian distance d between two points $p, q \in M$ is length of the shortest curve joining p to q:

$$d(p; q) = \inf \left\{ length(\gamma) : \ \gamma \text{ is horizontal curve that joins } p \text{ to } q \right\}.$$

In elementary Physics we refer to this infinum as the displacement. The curve over which this infinum is achieved is called a geodesic. It is trivially understood that such a distance

gives a natural cost function for the optimal control problems that exhibit structure of sub-Riemannian manifold. Moreover, in optimal control problems we are interested in finding the geodesics and the control input that materializes the geodesic.

Sub-Riemannian Geometry and Classical Mechanics

Sub-Riemannian geometry is the study of Sub-Riemannian manifolds that describe diverse physical systems but one particular example comes from classical mechanics where it is easiest to clarify the concepts associated with sub-Riemannian geometry. Classical mechanics bears innumerable examples of the systems whose velocities are constrained along subspaces of the tangent spaces thus conceding an obvious relationship with the distribution of a sub-Riemannian manifold.

The states of classical mechanical systems comprise of the generalized position and generalized velocity coordinates. Thus the phase space consists of the manifold M and the tangent bundle TM i.e., 'positions times speeds'. The derivatives of the position coordinates form the velocity coordinates thus imposing a natural constraint on the system dynamics. Hence the system trajectories are a subspace of the phase space and not all trajectories are allowed. Intuitively, one cannot vary speed without varying position and similarly, it is not possible to change position at zero velocity [40].

Historically Sub-Riemannian geometry was developed independently in Russia and France. In Russia, it is known as nonholonomic Riemannian geometry and in France it is known as Carnot geometry. Hence each source developed its own set of terminologies to describe various notions of the system. For example the sub-Riemannian distance is also known as Carnot-Caratheodory distance. For the interested reader, the term 'holonomic' was coined by German physicist Heinrich Rudolf Hertz. It literally means 'universal', 'integral', 'integrable', 'entire'. Nonholonomic therefore means 'non integrable' or 'not entire' suggesting that not entire tangent space is available for motion or the velocity constraints are integrated to position constraints [41].

Examples of Sub-Riemannian Manifolds in Physics

Sub-Riemannian geometry and the theory of nonholonomic systems have been the subject of active research since 1980s with its ramifications in control theory, classical mechanics, symplectic and contact geometry etc. Some of the most prominent geometers of the 20^{th} century have produced remarkable research papers on the subject and the number is increasing every passing day [41]. The underlying reason for such interest is due to the fact that optimal control of nonholonomic systems reduces to finding geodesics on the sub-Riemannian manifold. As such, sub-Riemannian manifold forms the phase space of a nonholonomic system. Phase space of any controllable dynamical system with number of vector fields smaller than the dimension of the phase space is a sub-Riemannian manifold. The state space of model of

steering of a car in example 3.3.1 is in fact a sub-Riemannian manifold and mathematically the car steering problem is a sub-Riemannian problem on group of motions of Euclidean plane. Sub-Riemannian problems arise in classical mechanics, economics, robotics, quantum mechanics, vision geometry and surprisingly in falling of cat [40]. Some of these problems are discussed to induce appreciation for sub-Riemannian geometry in control engineering.

Classical Mechanics - Robotics

Robotic systems are nonholonomic systems as we all know. Consider for example a two link planar robotic arm shown in figure 4.3 along with its configuration space. The links are of lengths $l_1$and l_2 respectively making angle θ_1 and θ_2 with the horizontal. The links can take angular velocities in the plane only and translational motion is not allowed. In that perspective the configuration space of the manipulator should be 1-sphere i.e., the circumference of the circle with radius l_1+l_2. However, we understand that two link pendulum can practically traverse a space bounded by two concentric circles with radii equal to l_1+l_2 and $l_1 - l_2$ respectively. Hence the velocity constraints do not integrate to position coordinates and therefore the system is nonholonomic i.e., non integrable and are studied under sub-Riemannian geometry.

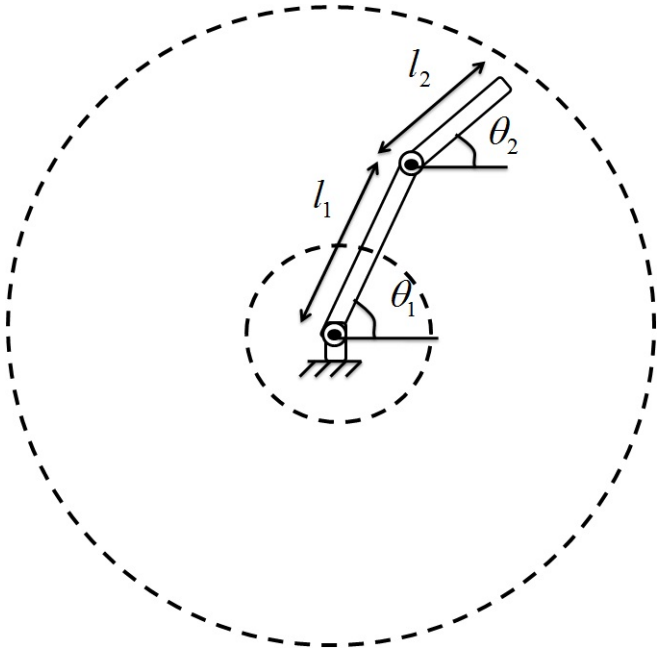

Figure 4.3: Two link planar robotic manipulator

Neuro Geometry of Vision

At the turn of the century important results were obtained in the neurophysiology of vision that clarified the functional mechanisms of the primary layer (V1) of the visual cortex of

human brain. The primary visual cortex contains different types of cells, including the so-called simple cells that perform the primary perception of visual information by the human brain. It was established that structure of the cortex is a sub-Riemannian manifold $M \equiv \mathbb{R}^2 \times \mathbb{S}^1$ [42].

This discovery proved to be vital in solving an important problem of restoring damaged or latent images called image inpainting. Researchers claim that if some portion of an image is missing or covered by another opaque object, then the brain attempts to complete the missing portion by minimizing some sub-Riemannian distance over the space of visual cells. In other words the brain computes a sub-Riemannian geodesic between the endpoints of the missing data. The problem has been posed as left-invariant sub-Riemannian problem on the group of motions of a Euclidean plane SE(2). A software program that computes the optimal curves was developed [43] and the results are promising. The researchers were able to restore gray scale images with missing visual data.

Quantum Mechanics

Quantum mechanics is the branch of Physics that deals with the motion of particles (atoms, molecules or system of particles with spin) on microscopic scale. The mechanics of such system is described by wave function $\psi : \mathbb{R} \to \mathbb{S}$. The function $\psi(t)$ is a vector containing statistical information about the quantum system. It is also called the quantum state or eigenstate equivalent to the concept of state in classical mechanics. The famous Schrödinger equation describes the time evolution of quantum mechanical system as:

$$ih\frac{\partial}{\partial t}\psi(t) = H_0\psi(t),$$

where i is imaginary unit, h is the Plank's constant and H_0 is the free Hamiltonian which is equivalent to the total energy of any given quantum mechanical system at any instant. It is possible to effect controls on a quantum system by applying external electric or magnetic fields. The control problem is to transfer the system form one eigenstate ψ_i to another eigenstate ψ_j by applying external stimuli $u_k, \quad k = 1, \ldots, n$, minimizing either the energy or time of transfer. This control problem is described on Heisenberg group which has a natural sub-Riemannian structure. Roughly speaking, the Riemannian metric is the energy transferred by the controls and the number of control vector fields is smaller than the dimension of the control system thereby taking structure of a distribution.

4.3 Sub-Riemannian Problem in General

Consider a driftless dynamical system on Riemannian manifold M of dimension n with metric g:

$$\dot{q} = \sum_{i=1}^{m} u_i(t) f_i(q), \qquad (u_1, \cdots, u_m) \in \mathbb{R}^m.$$

If $\Delta_q := span\{f_i(q)\} \subset T_qM \quad | \quad m < n$, then the problem of finding horizontal curves γ and the corresponding control u from some state q_0 to another arbitrary state q_1 such that sub-Riemannian length $l(q_0; q_1)$ is minimized is called a sub-Riemannian problem [38],[44]. It is easily seen that (M, Δ, g) form the sub-Riemannian manifold. The curve γ is called geodesic and the control u is called optimal control. Intuitively there exists a set of vector fields f_i whose values $\forall q \in M$ form a local orthonormal frame of the sub-Riemannian structure $(\Delta; g)$. The horizontal curves $\gamma(t) : I \subset \mathbb{R} \to M$ for $t \in I = [0, t_1]$ are the solutions of the following optimal control problem in M:

$$\begin{aligned} \dot{q} &= \sum_{i=1}^{m} u_i(t) f_i(q)), \qquad q \in M, \qquad (u_1, \cdots, u_m) \in \mathbb{R}^m, \\ q(0) &= q_0, \qquad q(t_1) = q_1, \\ l(\gamma) &= \int_0^{t_1} \sqrt{g\left(\dot{\gamma}(t), \dot{\gamma}(t)\right)} dt \to \min, \end{aligned} \tag{4.3.1}$$

where $l(\gamma)$ is the length of curve $\gamma : [0, t_1] \to M$ and suggests that we are looking for such a trajectory from q_0 to q_1 that minimizes this length.

Sub-Riemannian Length vs Energy Cost Functional

The sub-Riemannian length is seldom used as optimal cost functional in the sub-Riemannian problems. Instead, energy functional is minimized in sub-Riemannian problems. The energy of a parametrized curve $\gamma : [0, t_1] \to M$ is given as:

$$J(\gamma) = \frac{1}{2} \int_0^{t_1} g\left(\dot{\gamma}(t), \dot{\gamma}(t)\right) dt. \tag{4.3.2}$$

By Cauchy Schwarz inequality:

$$\left(\int_a^b f\left(t\right) g\left(t\right) \mathrm{d}t \right)^2 \leq \int_a^b \left(f\left(t\right)\right)^2 \mathrm{d}t \int_a^b \left(g\left(t\right)\right)^2 \mathrm{d}t.$$

Thus equation (4.3.1) may be written as:

$$\begin{aligned}
(l(\gamma))^2 &= \left(\int_0^{t_1} \sqrt{g\left(\dot{\gamma}(t),\dot{\gamma}(t)\right)}dt\right)^2, \\
&= \left(\int_0^{t_1} (1)\sqrt{g\left(\dot{\gamma}(t),\dot{\gamma}(t)\right)}dt\right)^2, \\
&\leq \left(\int_0^{t_1} (1)^2 dt\right).\left(\int_0^{t_1} \left(\sqrt{g\left(\dot{\gamma}(t),\dot{\gamma}(t)\right)}\right)^2 dt\right), \\
&= t_1 \int_0^{t_1} g\left(\dot{\gamma}(t),\dot{\gamma}(t)\right) dt. \qquad (4.3.3)
\end{aligned}$$

From equations (4.3.2) and (4.3.3) we have:

$$(l(\gamma))^2 \leq 2t_1 J(\gamma). \qquad (4.3.4)$$

Equation 4.3.4 essentially suggests that length minimization is equivalent to energy minimization problem. The functional J is more convenient than l since J is smooth and its extremals are automatically curves with constant velocity [15].

4.4 Sub-Riemannian Problem and Lie Groups

With sub-Riemannian geometry and its significance in perspective, it is now natural to inquire and establish the connection between sub-Riemannian problems and Lie groups. The craving for seeking such relationship ensues from the fact that this research is aimed at sub-Riemannian problem on Lie group SH(2). We have already established that Lie groups and geometry are important to the control theory. We have also established that sub-Riemannian problem is an optimal control problem with underlying manifold being sub-Riemannian. It only remains to establish, how sub-Riemannian problems are described on Lie groups which is dealt in the ensuing paragraphs motivated from [15].

As discussed earlier, a Lie group G is a smooth manifold that also has group structure. Associated to every Lie group is its Lie algebra $\mathcal{L}$ that locally determines the Lie group almost completely. Intuitively, Lie algebra is the tangent space at the identity of the Lie group i.e., $\mathcal{L} \equiv T_{Id}G$. For left invariant control systems $\dot{X} = XA, \quad X \in G, \quad A \in \mathcal{L}$, the tangent space at every $X \in G$ is the Lie algebra $\mathcal{L}$ transported to left by the group action on the Lie algebra basis A. Let us assume that vector field XA defines a distribution Δ_X and $\mathcal{L}$ is endowed with an invariant inner product $g = \langle\, .\, ,\, .\,\rangle$, then $\Delta_X \subset T_X G$ and g define a sub-Riemannian structure on Lie group G. For any Lipschitzian curve:

$$X : [0, t_1] \to M,$$

its Riemannian length is defined as integral of velocity:

$$l = \int_0^{t_1} \sqrt{\langle \dot{X}, \dot{X} \rangle} dt.$$

The sub-Riemannian problem is that given two arbitrary points $X_0, X_1 \in G$, find the curve in G that connects X_0 and X_1 such that the length l is minimized. The corresponding optimal control problem is as follows:

$$\begin{aligned} \dot{X} &= Xu, \quad X \in G, \quad u \in \mathcal{L}, \\ X(0) &= X_0, \quad X(t_1) = X_1, \quad X_0, X_1 \in G \quad \text{fixed}, \\ l(u) &= \int_0^{t_1} \sqrt{\langle u(t), u(t) \rangle} dt \rightarrow min. \end{aligned}$$

Hence left invariant control problem on a Lie group G is essentially a sub-Riemannian problem on G which in turn is an optimal control problem. To summarize:

- ▷ An optimal control problem is to find a curve γ on the manifold G such that some cost function is minimized.
- ▷ If the manifold G is a sub-Riemannian manifold then the sub-Riemannian length gives a natural cost function and the optimal control problem has sub-Riemannian structure.
- ▷ Additionally, if G has group structure and its Lie algebra is endowed with an invariant inner product, then the problem becomes sub-Riemannian problem on Lie group G.

4.5 Known Results - Sub-Riemannian Problem on Lie Groups

Past two decades have been a period of substantial activity in sub-Riemannian geometry from mathematics as well as control theory perspective. On theoretical front, it has been studied as basic model of metric spaces in the modern theory of analysis whereas, on applied side many problems have been modeled as sub-Riemannian problems in diverse areas such as control theory, economics, robotics, neurophysiology and image inpainting [37],[40].
Study of sub-Riemannian problems on various Lie groups via geometric control methods is a relatively new area of control theory. On macro scale sub-Riemannian problem is concerned with finding the optimal trajectories between states on manifold, but, on micro scale it deals with issues such as controllability, Liouville's integrability, existence, parametrization and qualitative analysis of extremal trajectories etc. Since PMP provides only necessary conditions for optimality, thus only candidate optimal trajectories known as extremal trajectories are obtained from it. Thereafter further analysis based on Maxwell points is performed to single out the extremal trajectories that are actually not optimal. The calculation of cut and conjugate loci becomes all the more important in this overall scheme of elimination procedure.

The scheme of analysis outlined above has been applied to various Lie groups. One of the most important Lie groups is Heisenberg group named after German theoretical physicist Werner Heisenberg renowned for his work in quantum mechanics. Three dimensional Heisenberg group arises in the description of one-dimensional quantum mechanical systems. Optimal control problem was considered on Heisenberg group in [45]. The motivation for this work were two physics problems i.e., Heisenberg flywheel with n point masses and the behavior of n non-interacting, non-relativistic charged masses under influence of a magnetic field. The authors applied PMP on the problem and calculated extremal trajectories that are solutions to the Hamiltonian system associated with the problem. Abnormal extremals (which do not depend on the cost functional) yield solutions that are geometrically irrelevant.

In a series of papers [46],[47] author considered the generalized Dido problem. The problem was posed as an optimal control problem in a 5-dimensional space. The control vector was 2-dimensional control with quadratic optimal cost functional. In the first paper, extremal trajectories of Dido's problem were obtained and parametrized by Jacobian elliptic functions. The second paper [47] was devoted to describing the reflection symmetries in the vertical subsystem of the Hamiltonian system. The Maxwell strata corresponding to reflection symmetries were calculated and upper bound on cut time was obtained in [46].

The rigorous scheme of analysis was then extended to group of motions of Euclidean plane that form the Lie group SE(2) [1]. More than hundred and fifty pages of analysis developed in [46],[47] were condensed to mere twenty three addressing all aspects handled in [46],[47]. The motions of Euclidean plane roughly represent a Reeds Shepp car [1] or a planar robot that is allowed to move forward and backward and turn only. Thus system has three dimensional vector field and two dimensional linear control input. Using Cauchy Schwarz inequality the author considers an optimal cost equivalent to the energy functional. Intuitively, such a system is controllable and the same is proved mathematically. The extremal trajectories are again parametrized by Jacobi elliptic functions. The strength of this scheme is that it has almost become an algorithm that has been applied to number of problems on various Lie groups. It is mathematically complex but leads to rigorous analysis and sound results.

In [48] the author addresses the question of cut and conjugate loci on Lie group SE(2). Author characterized the local and global optimality of extremal trajectories and calculated lower and upper bounds on the first conjugate time. It was proved that the cut time is equal to the first Maxwell time corresponding to the group of discrete symmetries of the exponential mapping. Another paper on SE(2) problem was dedicated to optimal synthesis[49].

The invariant Carnot Caratheodory metrics on Lie groups S^3, SL(2), SU(2), SO(3) and Lens spaces was studied in [50]. The authors calculated the geodesics and conjugate loci for all Lie groups. The cut loci were characterized globally and the expression for Carnot-Caratheodory distance was obtained. This was one of the most important work as it provided first explicit computation of the whole cut locus in sub-Riemannian geometry, which, at that time was known only for the trivial case of the Heisenberg group.

The problem of controllability of affine right-invariant systems on solvable Lie groups was addressed in [51]. After a span of fifteen years, the parametrization of extremal trajectories on SOLV^- was obtained in [52]. The extremal trajectories were yet again parametrized by Jacobi elliptic functions but the definition of elliptic coordinates was different than in [1]. The sub-Riemannian problem on SOL$V^-$is similar to and simpler than the sub-Riemannian problem on SH(2) which is the problem addressed in this research. However, the parametrization of sub-Riemannian geodesics obtained in [52] is far from complete.

Sub-Riemannian problem on Engel group was studied in [53]. It is essentially an optimal control problem described by 4-dimensional vector field and 2-dimensional linear control. It represents a nilpotent approximation to nonholonomic systems with 2-dimensional control in a 4-dimensional space. Intuitively, navigation of a mobile robot with trailer represents such a problem. Extremal trajectories are obtained and upper bound on cut time is calculated.

Interestingly, the problem of image inpainting (problem of recovering an image, some fragments of which are corrupted or hidden from observation) was taken up in [43]. The technique is based on completing damaged isophotes by sub-Riemannian length minimizers for the left-invariant sub-Riemannian problem on Lie group SE(2). An algorithm was proposed and a software *OptimalInpainting* was developed to restore the images. The results achieved from the software are promising and suggest the efficiency of technique. However, the mathematics and algorithm to restore colored images via sub-Riemannian length minimizers is yet to be realized.

Geometric control techniques have also been applied to physical systems. For example geometric tracking of a quadrotor was attempted in [54]. In this work authors expressed the dynamics of the quadrotor on special Euclidean group SE(3) as is customary in robotics. The authors achieved almost global asymptotic tracking of the attitude, position of the center of mass and the velocity of the center of mass of the quadrotor UAV. It was proved that system avoided singularities and could execute complex acrobatic maneuvers. The control was physically implemented and the quadrotor successfully performed a back flip.

In order to take advantage of computational power offered by computers, numerical solution techniques are a necessary part of the research. On that account, a new discipline has risen over past few years known as geometric computational mechanics [55] that provides numerical integration techniques for geometric control methods. In essence, these numerical integration methods preserve geometric properties of a dynamical system. This is important because the qualitative behavior of a dynamical system are determined by its geometric structure. In general the conventional integrators like Runge-Kutta, Euler method etc fail to provide reasonable computational results and the error keeps accumulating ultimately blowing up the result. With geometric integrators however, the results are consistent with the continuous time results and do not deteriorate over time. Among class of geometric integrators are included Lie group variational integrators developed in [56] for mechanical systems evolving on a Lie group. Their main advantage is that they preserve the geometric

structure and they exhibit good energy behavior over a long time period.

The number and variety of publications on geometric control theory and its applications are rising at ever increasing pace and therefore it is not possible to cover even a small number of such publications in this literature review. Hence, the literature review is by no means exhaustive or elaborate by design. Only important and most relevant results from past two decades have been included to highlight the activity that has engaged geometric control experts lately. Yet, we believe that the background of geometric control theory developed in this text is sufficient to explain the main concepts and motivate interested researchers towards a promising new direction of research.

4.6 Chapter Summary

Sub-Riemannian manifolds represent dynamical systems whose motion is restricted along horizontal spaces or which are subjected to nonholonomic constraints. Many mechanical systems are nonholonomic in nature and hence a geometric view of the problem can be certainly rewarding and insightful. Sub-Riemannian manifold is naturally endowed with a metric that represents distance between points and hence can be used as cost for optimal control problems. Hence PMP can be applied on the optimal control problem with cost being equivalent to the energy of the system.

In the succeeding chapter we formally define the Lie group SH(2) and state our research problem.

Chapter 5

Sub-Riemannian Problem on Group of Motions of Pseudo Euclidean Plane

In previous chapters the terminology and the concepts related to sub-Riemannian geometry were presented that completed the essential framework needed to define and explain the research problem. In this chapter we describe the problem statement in detail. We describe the notion of group of motions of pseudo Euclidean plane and describe the optimal control problem. Detailed description of the research objectives is also presented.

5.1 Motions

Definition 5.1.1. A transformation is a bijective mapping from a set A onto another set B [57].

Definition 5.1.2. A distance $d(x, y)$ preserving transformation $f : A \to B \quad | \quad d(x, y) = d(f(x), f(y))$ for any points $(x, y) \in A, \quad (f(x), f(y)) \in B$ is called an isometry. Distance d is any consistent notion of distance between elements of a set [57].

Definition 5.1.3. Motion is defined as isometry of a set onto itself [57].

We may think of motion as a transformation $f : A \to A$ such that distance between points $d(x, y) = d(f(x), f(y))$ is preserved. Motions represent the transformations of a rigid body where the intermediate distance between points on the body before and after transformation remains constant. Consider for example an aircraft under roll motion. The distance between wing tips before and after roll remains constant. Hence roll motion is an isometry (see Figure 5.1).

Properties of Motions

▷ The set of all motions is a group

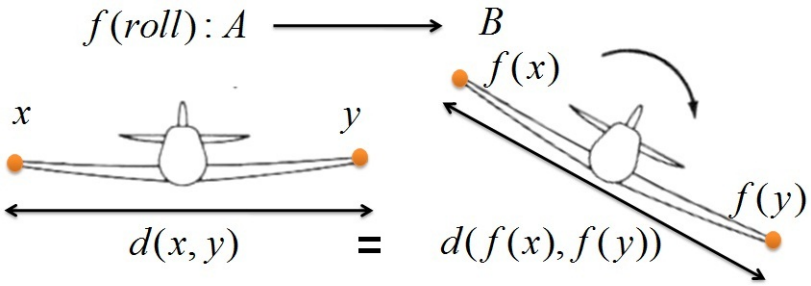

Figure 5.1: Aircraft under roll motion - Distance between wing tips is preserved

▷ Motions preserve the collinearity and noncollinearity of points, with the result that the image of a line is always a line

▷ Motions send a geometric object to a congruent geometric object

Types of Motion

1. Translation
2. Rotation
3. Reflection
4. Glide Reflection - Translation and reflection (see Figure 5.2).

Definition 5.1.4. Taken collectively, translations and rotations are called displacements, or rigid motions [57]. A motion $f : (x, y) \to (x', y')$ on Euclidean plane with translation a and b along x and y-axis respectively and rotation by angle θ is given as:

$$x' = x\cos\theta - y\sin\theta + a,$$
$$y' = x\sin\theta + y\cos\theta + b.$$

Displacements are the only motions which preserve the sense of every angle and are hence called direct transformations. Displacements form the Lie group called affine group $Aff(n)$.

5.2 Pseudo Euclidean Plane

Pseudo Euclidean space is a generalization of the Euclidean space. The generalization pertains to the possibility of having negative distances between points and a curved space. Mathematically, $(n + m)$ dimensional pseudo Euclidean space is denoted as F_m^{n+m} where F is a field over $\mathbb{R}$ and m is called the index. The space is endowed with a non-degenerate sign indefinite quadratic form $q(x) \quad \forall x \in \mathbb{R}^{n+m}$ [58]:

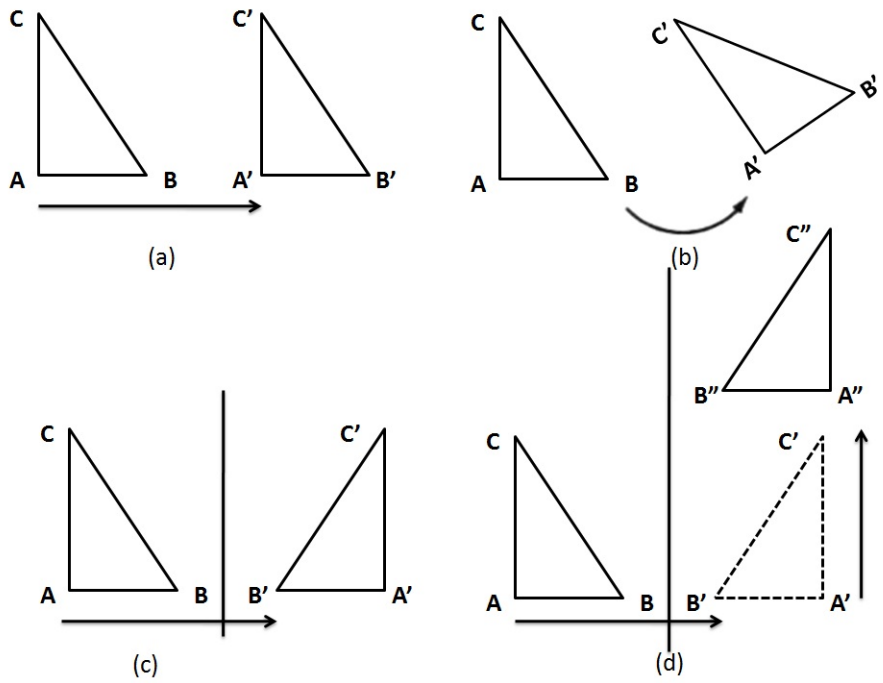

Figure 5.2: (a) Translation (b) Rotation (c) Reflection (d) Glide Reflection

$$q(x) = x^T Q x, \tag{5.2.1}$$

where Q is some real symmetric matrix. In terms of rectangular coordinates $(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m})$ quadratic form equation (5.2.1) is given as:

$$q(x) = (x_1^2 + \cdots + x_n^2) - \left(x_{n+1}^2 + \cdots + x_{n+m}^2\right).$$

Hence for pseudo Euclidean plane:

$$Q = \begin{pmatrix} I_n & 0 \\ 0 & -I_m \end{pmatrix}.$$

The pseudo Euclidean space also contains the notion of a polar form defined as:

$$\begin{aligned} \langle x, y\rangle &= x^T Q y, \\ &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n - (x_{n+1} y_{n+1} + \ldots + x_{n+m} y_{n+m}). \end{aligned} \tag{5.2.2}$$

Note that n-dimensional Euclidean space is a pseudo Euclidean space with index $m = 0$. Thus for Euclidean space $Q = I_n$ the quadratic form q and polar form $\langle x, y\rangle$ are given as:

$$\begin{aligned} q(x) &= x^T I_n x, \\ &= x_1^2 + \cdots + x_n^2. \end{aligned}$$

and,

$$\begin{aligned}\langle x, y\rangle &= x^T P y, \\ &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n. \end{aligned} \tag{5.2.3}$$

Looked this way, pseudo Euclidean space generalizes Euclidean space with $q(x)$ being equivalent to vector norm squared and polar form being the inner product of vectors. The distance r of a point from origin $\sqrt{q(x)}$ is Hermitian.

Pseudo Euclidean plane is principally two dimensional subspace F_1^{1+1} with $q(x) = x^2 - y^2$ where x and y represent the coordinate axes [58]. Rest of the discussion is now devoted to pseudo Euclidean plane except where mentioned explicitly. Distance r between two points $\mathbf{a}(a_1, a_2)$ and $\mathbf{b}(b_1, b_2)$ on F_1^{1+1} is given by the inner product equation (5.2.2):

$$r^2 \equiv \langle \mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{b}\rangle = (a_1 - b_1)^2 - (a_2 - b_2)^2.$$

Unlike the Euclidean plane, the distance r in pseudo Euclidean plane is Hermitian and can also be zero between two distinct points $\mathbf{a}(a_1, a_2)$ and $\mathbf{b}(b_1, b_2)$ provided $a_1 - b_1 = \pm(a_2 - b_2)$. Geometrically, unit hyperbola $x^2 - y^2 = 1$ shown in Figure 5.3 represents pseudo Euclidean plane F_1^{1+1} which is segregated into four distinct sectors by the asymptotic lines $x = y$ and $x = -y$. The sectors of the plane are conventionally known as Right (RS), Up (US), Left (LS) and Down (DS) sectors. The motions of pseudo Euclidean plane considered in this research problem are sector preserving maps on RS where $-x < y < x$ and distance between points $r > 0$.

The hyperbolic functions, gyrovectors and split complex numbers are pseudo Euclidean analogs of trigonometric functions, vectors and complex numbers in the Euclidean plane. Hence the points on RS of the pseudo Euclidean plane are transformed into polar coordinates by hyperbolic functions as follows [59],[60]:

$$\begin{aligned} a_1 &= r\cosh\varphi, \\ a_2 &= r\sinh\varphi, \end{aligned}$$

where $r \in \mathbb{R}^+$ is the length and $\varphi \in \mathbb{R}$ is the hyperbolic angle of rotation of the gyrovector. The split complex numbers also called hyperbolic/hypercomplex/perplex numbers represent the points on pseudo Euclidean plane in similar way as the complex number define points on the Euclidean plane. The number system was invented by Sophus Lie. It is defined as:

$$\mathbf{a} = a_1 + i a_2, \tag{5.2.4}$$

where $i^2 = +1$. The conjugate of split complex number is defined as $\bar{\mathbf{a}} = a_1 - i a_2$. Hence the quadratic form $q(.)$ on pseudo Euclidean plane is alternatively defined as:

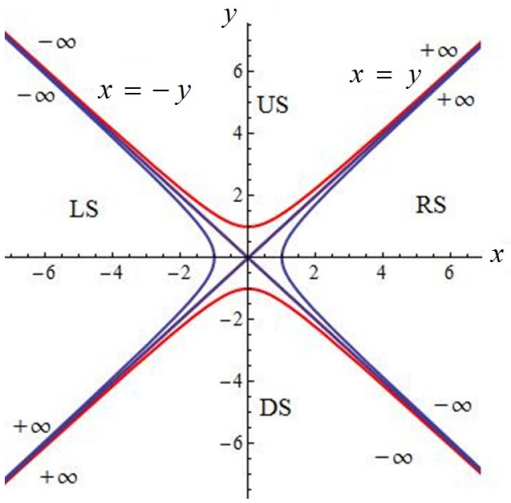

Figure 5.3: Pseudo Euclidean plane represented by unit hyperbola

$$\begin{aligned} q &= \mathbf{a}\bar{\mathbf{a}}, \\ &= (a_1 + ia_2)(a_1 - ia_2), \\ &= a_1^2 - i^2 a_2^2, \\ &= a_1^2 - a_2^2. \end{aligned} \tag{5.2.5}$$

The scalar product and the properties of hypercomplex numbers allow to state suitable axioms and to give the pseudo-Euclidean plane the structure of a vector space.

5.3 Group SH(2) of Motions of Pseudo Euclidean Plane

We now describe the motions of the pseudo Euclidean plane that comprise of translations and hyperbolic rotations. Subsequent exposition is based on relevant discussion in [60]. Specifically we consider displacements (see definition 5.1.4) of points in the pseudo Euclidean plane i.e., non homogeneous, linear maps of points of pseudo Euclidean plane that preserve distance, orientation and sector of points being transformed. Transformation of a point $\mathbf{a}(a_1, a_2)$ into another point $\mathbf{b}(b_1, b_2)$ by translating it a distance x and y along x and y-axis in RS (see figure 5.3) and rotated by angle z is given as:

$$\begin{aligned} b_1 &= a_1 \cosh z + a_2 \sinh z + x, \\ b_2 &= a_1 \sinh z + a_2 \cosh z + y, \end{aligned}$$

where $x, y, z \in \mathbb{R}$. The transformation $m : \mathbf{a} \to \mathbf{b}$ is called the motion m of the pseudo Euclidean plane. We see that the motion m of pseudo Euclidean plane is completely

parametrized by $x, y, z \in \mathbb{R}$. Composition of two motions $m_1(x_1, y_1, z_1)$ and $m_2(x_2, y_2, z_2)$ is another motion $m_3(x_3, y_3, z_3)$ given as:

$$m_3(x_3, y_3, z_3) = m_1(x_1, y_1, z_1).m_2(x_2, y_2, z_2),$$

where,

$$\begin{aligned} x_3 &= x_2 \cosh z_1 + y_2 \sinh z_1 + x_1, \\ y_3 &= x_2 \sinh z_1 + y_2 \cosh z_1 + y_1, \\ z_3 &= z_1 + z_2. \end{aligned}$$

The identity motion m_{Id} is given by $x = y = z = 0$, and inverse of a motion $m(x, y, z)$ is given by $m^{-1}(x^1, y^1, z^1)$ where,

$$\begin{aligned} x^1 &= -x \cosh z + y \sinh z, \\ y^1 &= x \sinh z - y \cosh z, \\ z^1 &= -z. \end{aligned}$$

The composition of motion m with m^{-1} is given as $m^{-1}m = m_{Id}$. Note that motions of pseudo Euclidean plane have a group structure with composition as the group operation.

5.4 Homogeneous Coordinates and Matrix Representation of Group of Motions of Pseudo Euclidean Plane

Groups and group operations can be represented by matrices. In order to represent the motions of pseudo Euclidean plane we need to choose a plane to use for which apparently $z = 0$ is the most suitable choice. However, it turns out that $z = 1$ is a better choice which means that the plane is situated at height of $z = 1$ in the 3-dimensional space. Thus every point in the plane has so called homogeneous coordinate of the form [60]:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

The motions of the plane can be represented by 3×3 matrices of the form:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{pmatrix} \tag{5.4.1}$$

The motion of a point now becomes the matrix multiplication with the column vector representing the point.

5.5 Lie Group and Lie Algebra Representation

Specifically, for pseudo Euclidean plane the transformation matrix is given as [60]:

$$M = \mathrm{SH}(2) = \left\{ \begin{pmatrix} \cosh z & \sinh z & x \\ \sinh z & \cosh z & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}, \tag{5.5.1}$$

where SH(2) stands for special hyperbolic group of dimension 2. The group SH(2) also has a smooth manifold structure. Note that it represents a configuration space with 3 degrees of freedom i.e., two translational degrees of freedom and one rotation degree of freedom. Hence, the configuration manifold is three dimensional or we need three independent coordinates to define configuration of a rigid body on the manifold. Being a smooth manifold and a group, SH(2) obviously is a Lie group with matrix multiplication as the group operation. In keeping with the motivation behind Lie groups, they represent continuous transformations of the points on pseudo Euclidean plane.

The Lie group SH(2) comprises three basis one-parameter subgroups given as:

$$w_1(t) = \begin{pmatrix} \cosh t & \sinh t & 0 \\ \sinh t & \cosh t & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w_2(t) = \begin{pmatrix} 1 & 0 & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w_3(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$$

The basis subgroups w_1, w_2, w_3 represent rotation by angle t and translation by distance t along x and y-axis respectively. When applied individually to a point, they result in a corresponding motion only i.e., rotation or translation along either axis of the plane. The corresponding basis for Lie algebra are the tangent matrices $A_i = \frac{dw_i(t)}{dt} \mid_{t=0}$ to the subgroups of Lie group SH(2). A_i are given as:

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.5.2}$$

The Lie algebra is thus:

$$\mathcal{L} = T_{Id}M = \mathrm{sh}(2) = \mathrm{span}\left\{A_1, A_2, A_3\right\}.$$

The multiplication rule for $\mathcal{L}$ is $[A, B] = AB - BA$. Therefore, the Lie bracket for sh(2) is given as $[A_1, A_2] = A_3$, $[A_1, A_3] = A_2$ and $[A_2, A_3] = 0$ [2].

5.6 Sub-Riemannian Problem on SH(2)

Consider the following driftless control system on SH(2):

$$\begin{aligned} \dot{q} &= f_1(q)\,u_1 + f_2(q)\,u_2,\ q \in M = \mathrm{SH}(2),\ (u_1, u_2) \in \mathbb{R}^2, && (5.6.1)\\ q(0) &= Id, \qquad q(t_1) = q_1, && (5.6.2)\\ l &= \int_0^{t_1} \sqrt{u_1^2 + u_2^2}\, dt \to \min, && (5.6.3)\\ f_1(q) &= qA_2, \qquad f_2(q) = qA_1. && (5.6.4) \end{aligned}$$

Here equation (5.6.1) represents the dynamical system with control inputs u_i and control distribution $\Delta = \mathrm{span}\{f_1, f_2\}$. The states $q(0)$ and $q(t_1)$ represent the initial state at $t = 0$ and final state to be reached at time t_1. In equation (5.6.3) l is the sub-Riemannian length functional to be minimized. Canonical frame on M in terms of [2] is given as:

$$f_1(q),\ f_2(q),\ f_0(q) = qA_3,$$

$$[f_1, f_0] = 0, \qquad [f_2, f_0] = f_1, \qquad [f_2, f_1] = f_0. \tag{5.6.5}$$

By [2], the sub-Riemannian structure:

$$(M, \Delta, g), \qquad \Delta = \mathrm{span}\{f_1, f_2\}, \qquad g(f_i, f_j) = \delta_{ij},$$

is unique upto rescaling, left invariant contact sub-Riemannian structure on SH(2). Here δ_{ij} is the Kronecker delta.

The system may be visualized in terms of a planar robot or a Reeds-Shepp car moving on a pseudo Euclidean plane. The car has three degrees of freedom on 2 dimensional plane i.e., translation along x and y axis and rotation about a vertical axis passing through its center of mass. The car's center of mass lies at coordinates (x, y) and the angle z is its orientation in the plane. Given the initial and final states, the objective is to calculate a horizontal curve $\gamma \subset \mathrm{SH}(2)$ between the q_0 and q_1 such that the sub-Riemannian length functional given in equation (5.6.3) is minimized. The admissible trajectories are assumed Lipschitz and admissible control inputs u_i are assumed bounded.

5.7 Motivation

> The principle of relativity corresponds to the hypothesis that the kinematic space is a space of constant negative curvature, the space of Lobachevsky and Bolyai. The value of the radius of curvature is the speed of light.
>
> Borel 1913

Pseudo Euclidean plane is more commonly known as hyperbolic plane. However, we intentionally deferred calling it so until now where we present the reasons why pseudo Euclidean plane is called hyperbolic plane. As argued earlier, Euclidean geometry reined the mathematical world and classical mechanics until 19^{th} century. Although accepted almost as the only geometry describing the universe, the parallel postulate of Euclidean geometry was viewed with suspicion and remained subject of research by geometers following Euclid. In an attempt to disprove the parallel postulate, Nikolai Lobachevsky in 1829 and Janos Bolyai in 1832 independently discovered Non-Euclidean geometry also known as hyperbolic geometry [61]. Until E. Beltarami, Hyperbolic Geometry remained cutoff from the rest of the mathematics who in 1868 proved that two dimensional non-Euclidean geometry was study of surfaces with constant negative curvature.

Discovery of hyperbolic geometry paved the way for the formulation of Special Theory of relativity by Einstein in 1905 [62]. The central concept of Einstein's seminal paper of 1905 was velocity addition law that satisfies hyperbolic parallelogram addition of velocities and which is modeled on Bolyai-Lobachevsky's hyperbolic geometry. In 1907, Hermann Minkowski reformulated his former student Albert Einstein's special theory of relativity. Minkowski introduced the concept of four dimensional space, known as "Minkowski space-time", with 3 ordinary dimensions of space and another intermingled dimension of time. Minkowskian space-time thus presented a mathematical setting in which Einstein's theory of relativity and Lorentz geometry could be formulated.

The Minkowskian space-time geometry is naturally linked to four dimensional pseudo Euclidean space with index $m = 3$ and traditionally different authors have used the phrase hyperbolic plane to refer both to (Bolyai-Lobachevskian) hyperbolic geometry and Minkowski geometry but these are two different geometries. Space-time is described by Minkowski space, but the velocity space is described by hyperbolic geometry. Hence, it is out of this misnomer and because the hyperbolic numbers represent points on pseudo Euclidean plane, it is referred to as hyperbolic plane. We will follow the convention and use the hyperbolic plane and pseudo Euclidean plane interchangeably to refer to the same plane i.e., plane of hyperbolic numbers.

A pseudo Euclidean plane represents Mikowskian space-time of two dimensions with one spatial variable and one temporal variable. Thus the isotropic cone formed by lines $x = y$ and $x = -y$ consists of so called light-like vectors. When $q(x) < 0$, the vectors are called space like vectors and when $q(x) > 0$, the vectors are called time like vectors. Introduction of hyperbolic geometry and gyrovectors had a profound impact on Einstein's theory of relativity and hence our understanding of the laws of universe. It allowed vector algebra to be introduced in relativistic mechanics and hence made further development in relativistic mechanics possible. Hence its study can possibly give insight into how nature works. From perspective of relativistic mechanics, defining and solving sub-Riemannian problem on group SH(2) can possibly lead to better understanding of the theory of relativity and enhancement of our understanding of the universe.

Sub-Riemmnian problems are viewed as the optimal control problems of nonholonomic systems with the optimal cost being the sub-Riemannian length. Sub-Riemannian problem on Lie group SH(2) represents the optimal control of a unicycle moving on a hyperbolic plane. The mathematical model of a unicycle theoretically is the basis for modelling a large class of nonholonomic systems in robotics e.g., a differential drive robot, UAVs etc [63]. Traditionally the robots are considered to be moving the Euclidean space ignoring the fact that practically all physical spaces are combinations of hyperbolic and spherical surfaces i.e., there are no flat surfaces in the real environment. Consequently, the control and motion planning algorithms solved for Euclidean space are less effective in the physical world and the use of robots is restricted to controlled or laboratory environment. Since this work considers the motion of such robots on hyperbolic surfaces, the setting and therefore the results are more realistic. Therefore, we expect the results of this research to have potential impact in the field of robotics.

This research involves the tools of geometric control theory which amongst the various notions of control is a realtively newer and less developed. It is inherently a nonlinear optimal control design technique based on PMP. The other competing notion of optimal control for nonlinear systems is the HJB equations. However, as noted earlier, the limitations of HJB equations and dynamic programming make it far more complicated for control of practical systems. In words of Pontryagin himself, “It is interesting to note that the Bellman approach does not actually constitute a mathematical solution, but rather a "good heuristic" (in Pontryagin’s words). The maximum principle on the other hand is an actual mathematical solution”. Hence, the analysis techniques developed and results obtained in this research can potentially have far reaching implications in the design of actual robots moving in real environments.

In [2] provided a classification of all sub-Riemannian structures on three dimensional Lie groups in terms of the basic differential invariants. Hence, sub-Riemannian problem on group SH(2) is very important in the entire study of three dimensional Lie groups.

Research Objectives

The overall research objectives of any sub-Riemannian problem are following:

1. Controllability, integrability and existence of optimal trajectories of the dynamical system, [4],[3].
2. Obtain complete parametrization of extremal trajectories [3].
3. Description of symmetries of the system and the corresponding Maxwell sets [3], [5].
4. Characterize conjugate loci [5].

5. Geometric view of extremal trajectories and Maxwell Strata through 3D plots sub-Riemannian spheres [5].

6. Description of cut loci and global optimality analysis [49], [64].

7. Description of the global structure of the exponential mapping and the optimal synthesis.

8. Geometric analysis of the cut locus and conjugate locus through 3D plot of sub-Riemannian caustic

The idea is to apply techniques of geometric control theory [20, 15, 1, 48, 49]. Note that objectives 1–4 have been addressed in this research thesis whereas objectives 5–8 were stated as part of future work and are beyond the scope of this work.

5.8 Explanation of Research Objectives

Controllability and Integrability of the Dynamical System

The objective is to prove ODEs given by (5.6.1) describing dynamical system defined in equations (5.6.1–5.6.4) are explicitly integrable in terms of some mathematical functions and that the system is globally controllable. Proof of controllability ensures that any two points on the manifold can be joined by a real analytic curve. Proof of integrability is also extremely important before attempting to parametrize the extremal trajectories which is the natural next objective of research on SH(2). Existence of optimal trajectory guarantees that a sub-Riemannian length minimizing geodesic exists between any two points on the manifold M.

Parametrization of Extremal Trajectories

Pontryagin's Maximum Principle yields necessary but not sufficient conditions for optimal trajectory. The trajectories that satisfy the Maximum Principle are called **extremal trajectories**, and are candidate optimal only. The Maximum Principle provides a compact geometrical description of the extremal trajectories, and thus gives us a tool for enumerating and exploring optimal trajectories. Parametrization pertains to obtaining parametric equations of extremal trajectories in minimum number of parameters such that at any value of these parameters the equations of extremal trajectories give a point in configuration space. It can also be termed as the integration of nonlinnear differential equations to obtai parametric equations of the configuration variables of the system. In order to apply second order optimality conditions it is necessary to obtain a suitable parametrization. Parametrization provides a convenient mathematical description of candidate optimal trajectories and eases or makes further optimality analysis tractable.

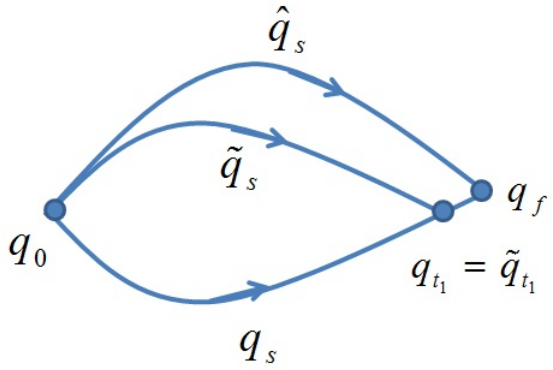

Figure 5.4: Non-optimality of the geodesic q_s after the Maxwell point q_{t_1}

Description of Maxwell Strata

The locus of the intersection points of extremal trajectories with equal sub-Riemannian lengths is called Maxwell Set [46]. The Maxwell set is closely related to optimality of extremal trajectories. It was proved by S. Jacquet that an extremal trajectory cannot be optimal after a Maxwell point [65]. Let q_s and $\tilde{q}_s$ be two distinct extremal trajectories i.e., $q_s \neq \tilde{q}_s$, $s \in [0, t_1]$ (see Figure5.4). If $q_{t_1} = \tilde{q}_{t_1}$, then for any $t_2 > t_1$ the extremal trajectories q_s and $\tilde{q}_s$, $s \in [0, t_2]$ are not optimal.
A system with symmetries has obvious Maxwell points for extremal trajectories that are reflections or rotations symmetries of each other and cross one another at some point. Such trajectories being symmetry of each other obviously have equal sub-Riemannian length. Hence the idea is to shrink the extremal trajectories set by eliminating the ones that happen to be symmetries of each other and form Maxwell points [46].

Characterize Conjugate Loci and sub-Riemannian Spheres

Definition 5.8.1. Conjugate Locus

The set of points where extremal trajectories lose local optimality (i.e., optimality with respect to sufficiently close extremal trajectories) [48].

Definition 5.8.2. Sub-Riemannian Sphere

Sub-Riemannian Sphere with center q_0 of radius R is the set of points q_1 located at the sub-Riemannian distance R from the point q_0 [48].
Thus the objective of this research problem is to characterize and analyze sub-Riemannian geodesics, the cut and conjugate loci and sub-Riemannian sphere on the group SH(2). Detailed discussion about conjugate loci is given in chapter 9.

Description of Cut Loci

Definition 5.8.3. Cut Locus

The set of points where extremal trajectories lose global optimality (i.e., optimality with respect to all other extremal trajectories) [48].

This analysis pertains to further eliminating the extremal trajectories that seize to be optimal for some reason. In this part of the analysis the cut time i.e., the time of loss of global optimality is proved analytically. The cut time is shown to be equal to the first Maxwell time corresponding to the group of discrete symmetries of the exponential mapping.

Global Structure of the Exponential Mapping and the Optimal Synthesis

An exponential map is a mapping from Lie Algebra to the corresponding Lie Group. In terms of linear systems it is the ordinary exp function that is obtained by integrating the differential equations (Lie Algebra) and results in the reachable set of the system that lies in the Lie Group. The exponential map for generalized manifolds and nonlinear systems is not the trivial exp function. The objective of this research problem is therefore to describe the exponential mapping $\gamma : sh(2) \longrightarrow \mathrm{SH}(2)$, where sh(2) is the Lie Algebra of the group SH(2). The solution will also give a generalized framework through which the optimal control problem will be synthesized and solved on SH(2).

5.9 Chapter Summary

The Lie group SH(2) is group of motions of hyperbolic plane that is extremely important in theory of relativity and sub-Riemannian problem on SH(2) is certainly an important problem. It is equivalent to considering a planar robot moving on a surface with constant negative curvature. Since the real world surfaces are seldom flat, hence this is an important nonholonomic problem.

In the next chapter we consider the controllability and integrability analysis of the research problem (5.6.1)–(5.6.4).

Chapter 6

Controllability and Integrability of the Dynamical System

We now consider three most important problems in relation to the sub-Riemannian problem on Lie group SH(2) i.e., the problem of controllability, the existence of optimal controls and possibility to integrate the dynamical system and obtain a close form solution in terms of elementary mathematical functions. To prove integrability of the dynamical system, we convert the matrix representation of the Lie group SH(2) into vector form using Wei-Norman representation. Controllability and integrability analysis paves the way for parametrization of extremal trajectories.

6.1 Left Invariance

Claim 6.1.1. The sub-Riemannian problem defined in equations (5.6.1)–(5.6.4) is a left invariant problem on Lie Group SH(2).

Proof. A left invariant problem is of the form:

$$\dot{q} = qAu \quad | \quad A \in \mathcal{L}, \quad q \in G, \quad u \in U \subset \mathbb{R}. \tag{6.1.1}$$

Defintion of a left invariant vector field is given in (2.15.1). A left invariant control system is an aribtrary set of vector fields [15]. It can be readily seen that the vector field qA is left invariant. Intuitively, a left invariant vector field means that multiplying the Lie algebra $\mathcal{L} = T_{\mathrm{Id}}G$ with the group element on the left transports the tangent space at the identity of the Lie group G to point $q \in G$. Hence the tangent space on the entire group G is the transported version of $\mathcal{L}$. Note that A is a matrix representation of $\mathcal{L}$. We now prove that the control system (5.6.1) is also left invariant. Consider (5.5.1):

$$q = \begin{pmatrix} \cosh z & \sinh z & x \\ \sinh z & \cosh z & y \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{SH}(2). \tag{6.1.2}$$

Taking time derivative of q:

$$\begin{aligned}
\dot{q} &= \begin{pmatrix} \sinh z\,\dot{z} & \cosh z\,\dot{z} & \dot{x} \\ \cosh z\,\dot{z} & \sinh z\,\dot{z} & \dot{y} \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} u_2 \sinh z & u_2 \cosh z & u_1 \cosh z \\ u_2 \cosh z & u_2 \sinh z & u_1 \sinh z \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & \cosh z \\ 0 & 0 & \sinh z \\ 0 & 0 & 0 \end{pmatrix} u_1 + \begin{pmatrix} \sinh z & \cosh z & 0 \\ \cosh z & \sinh z & 0 \\ 0 & 0 & 0 \end{pmatrix} u_2 \\
&= qA_2u_1 + qA_1u_2. \qquad (6.1.3)
\end{aligned}$$

Hence the sub-Riemannian problem (5.6.1)–(5.6.4) is a left invariant problem on the Lie Group SH(2). □

Left invariance (or alternatively right invariance) has important implications. Essentially, the tangent space spanned by the vector fields is the same as that at the identity of the Lie group. Hence the parametrization of trajectories can be obtained at the group identity i.e., $x(0) = y(0) = z(0) = 0$ and be transported to any point $q \in G$ by the group action. Hence for integration of dynamical system initial conditions $x(0) = y(0) = z(0) = 0$ suffice.

6.2 Controllability

Theorem 6.2.1. *The dynamical system (5.6.1) is completely controllable. Moreover, the optimal control problem (5.6.1)–(5.6.4) has solutions.*

Rashevskii Chow's Theorem

Before we give the proof of this theorem we state the famous theorem proved independently by Petr Konstanovich Rashevskii in 1938 [66] and W. L. Chow (China) in 1939 [67]. It states that:

Let Δ be a differential system or distribution totally nonholonomic at each point of a smooth manifold M, then each two points $X, Y \in M$ may be connected by an admissible piecewise smooth curve of finite length.

OR

If Lie algebra of the control distribution $\mathcal{L}_q\Delta = T_qM \quad \forall \quad q \in M$ and if M is connected, then each orbit is equal to the whole manifold M.

OR

For a connected manifold and corresponding bracket generating control distribution, the system is completely controllable.
Intuitively, it means to say that if the configuration manifold for a dynamical system is connected (i.e., there are no voids in the state space) and the number of independent vector fields in the control distribution is smaller than the dimension of the Lie algebra, but, we can find sufficient number of independent Lie brackets, then essentially we can traverse the entire tangent space and hence reach every point of the configuration manifold. Rashevskii–Chow theorem is one of the corner stones of Geometric control theory.

Proof. The control distribution $\Delta = \text{span}\{f_1(q)\ , f_2(q)\}$ is full rank because from (5.6.5) $[f_1(q)\ , f_2(q)] = -[f_2(q), f_1(q)] = -f_0(q)$. It satisfies *Hörmander condition* or the bracket generating condition. Hence the Lie algebra of the distribution $\mathcal{L}_q\Delta$ is given as:

$$\mathcal{L}_q\Delta = \text{span}\{f_1(q), f_2(q), -f_0(q)\} = T_q\text{SH}(2) \qquad \forall q \in M.$$

The manifold SH(2) is connected and the system is bracket generating. Hence the system satisfies Rashevskii-Chow's Theorem and is therefore completely controllable.
Existence of optimal trajectories for problem given by equations (5.6.1-5.6.4) follows from Filippov's theorem [20]. □

6.3 Local Representation of a Control System

Having dealt with the issue of controllability, we turn towards another important question i.e., how to construct specific trajectories? In other words this pertains to finding integrating the nonlinear differential equations of the control system (5.6.1) defined on the matrix Lie group. The properties like controllability of a control system can be established directly from the algebraic and geometric properties of its global description:

$$\dot{q} = q\sum_{i=1}^{m} A_i(t)u_i(t), \quad q \in G, \quad A_i \in \mathcal{L}, \quad \bigcup u_i \in \mathbb{R}^m, \tag{6.3.1}$$

where G is an n-dimensional Lie group, A_i are the basis of Lie algebra $\mathcal{L}$ and u_i are the open loop control variables. Global representation however is not the most convenient representation for other purposes such as computing the control laws for which we need to resort to local representations of (6.3.1) when we compute the actual control laws.

There are two different ways to construct trajectories, using two types of local representations:

- ▷ Magnus representation - Canonical coordinates of the first kind [68]
- ▷ Wei-Norman representation - Canonical coordinates of the second kind [69]

Such local representation expresses the control system in terms of coordinates $x_i, \quad i = 1, \ldots, n$:

$$\dot{x} = f_1(x)u_1 + f_2(x)u_2 + \ldots + f_m(x)u_m, \quad x \in \mathbb{R}^n, \quad (u_1, u_2) \in \mathbb{R}^m, \tag{6.3.2}$$

where f_i are the vector fields. The local representation (6.3.2) is only valid in a neighborhood N of $Id \in G$.

As is evident, a control engineer is more accustomed to dealing with control system of the form (6.3.2) which involves vectors instead of matrices. Note that (6.3.2) is representation of a drift free system but systems with drift are also represented in a similar way.

Wei-Norman Representation

We state the following theorem without proof by J. Wei and E. Norman (1964) that expresses the dynamical system (6.3.1) in canonical coordinates of second kind.

Theorem 6.3.1. *(Wei and Norman [69]) Consider the driftless control system (6.3.1) and its solution $q(t), \quad t \geq 0$. Then in a neighborhood of $t = 0$ the solution may be expressed in the form:*

$$q = e^{x_1 A_1} e^{x_2 A_2} \ldots e^{x_n A_n}. \tag{6.3.3}$$

The coordinate functions $x_i(t)$ evolve according to:

$$\begin{pmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix} = F(x_1, \ldots, x_n) \begin{pmatrix} u_1(t) \\ \vdots \\ u_n(t) \end{pmatrix}, \tag{6.3.4}$$

where F is analytic in coordinates x_i and depends only on the structure of the Lie algebra $\mathcal{L}$. Moreover, if the Lie algebra is solvable, then there exists a basis and an ordering of this basis for which the representation (6.3.4) is global and the coordinates x_i can be computed by quadratures.

For obvious reasons, this is also called product of exponential representation. This representation also has following properties for the control system (5.6.1) [70]:

- ▷ coordinates have a large, well defined region of validity;
- ▷ representation is in terms of explicit functions;
- ▷ representation reflects structural properties of (6.3.1);
- ▷ representation has good numerical properties with respect to integration;
- ▷ coordinates have physical interpretation.

Let us now construct the Wei-Norman representation for the left invariant control system on the Lie group SH(2) (5.6.1). Recall that our control system is given as (6.1.3):

$$\dot{q} = qA_2u_1 + qA_1u_2, \tag{6.3.5}$$

where,

$$q = \begin{pmatrix} \cosh z & \sinh z & x \\ \sinh z & \cosh z & y \\ 0 & 0 & 1 \end{pmatrix}.$$

Recall also that basis A_1, A_2, A_3 of Lie algebra sh(2) are given as:

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.3.6}$$

According to Wei-Norman formula (6.3.3) the local solution to the control system (6.3.5) may be written as:

$$\begin{aligned} q &= e^{x_1A_1}e^{x_2A_2}e^{x_3A_3}, \\ \dot{q} &= e^{x_1A_1}\dot{x}_1A_1e^{x_2A_2}e^{x_3A_3} + e^{x_1A_1}e^{x_2A_2}\dot{x}_2A_2e^{x_3A_3} + e^{x_1A_1}e^{x_2A_2}e^{x_3A_3}\dot{x}_3A_3 \\ &= e^{x_1A_1}e^{x_2A_2}e^{x_3A_3}e^{-x_3A_3}e^{-x_2A_2}\dot{x}_1A_1e^{x_2A_2}e^{x_3A_3} + e^{x_1A_1}e^{x_2A_2}e^{x_3A_3}e^{-x_3A_3}\dot{x}_2A_2e^{x_3A_3} \\ &\quad + e^{x_1A_1}e^{x_2A_2}e^{x_3A_3}\dot{x}_3A_3 \\ &= q(\dot{x}_1e^{-x_3A_3}e^{-x_2A_2}A_1e^{x_2A_2}e^{x_3A_3} + \dot{x}_2e^{-x_3A_3}A_2e^{x_3A_3} + \dot{x}_3A_3). \end{aligned}$$

Using Baker-Capmbell-Hausdorf formula $e^ABe^{-A} = e^{ad_A}B$ where $ad_A = [A, B]$, last equation may be written as:

$$\dot{q} = q\left(\dot{x}_1e^{ad_{-x_3A_3}}e^{ad_{-x_2A_2}}A_1 + \dot{x}_2e^{ad_{-x_3A_3}}A_2 + \dot{x}_3A_3\right), \tag{6.3.7}$$

Now using the Lie bracket of section 5.5 $e^{ad_{-x_3A_3}}A_2 = A_2 - x_3[A_3, A_2] = A_2$ and $e^{ad_{-x_2A_2}}A_1 = A_1 - x_2[A_2, A_1] = A_1 + x_2A_3$. Similarly,

$$\begin{aligned} e^{ad_{-x_3A_3}}e^{ad_{-x_2A_2}}A_1 &= e^{ad_{-x_3A_3}}\left(e^{ad_{-x_2A_2}}A_1\right), \\ &= e^{ad_{-x_3A_3}}\left(A_1 + x_2A_3\right) \\ &= e^{ad_{-x_3A_3}}A_1 + x_2e^{ad_{-x_3A_3}}A_3 \\ &= A_1 - x_3[A_3, A_1] + x_2\left(A_3 - x_3[A_3, A_3]\right) \\ &= A_1 + x_3A_2 + x_2A_3. \end{aligned}$$

Substituting formulas of e^{ad_-} in (6.3.7) we have:

$$\begin{aligned} \dot{q} &= q\left(\dot{x}_1(A_1 + x_3A_2 + x_2A_3) + \dot{x}_2A_2 + \dot{x}_3A_3\right) \\ &= q\left(\dot{x}_1(A_1 + x_3A_2 + x_2A_3) + \dot{x}_2A_2 + \dot{x}_3A_3\right) \\ &= q\dot{x}_1A_1 + q(\dot{x}_1x_3 + \dot{x}_2)A_2 + q(\dot{x}_1x_2 + \dot{x}_3)A_3. \end{aligned} \tag{6.3.8}$$

Comparing now coefficients of A_1, A_2, A_3 in (6.3.5) and (6.3.8) we have:

$$\begin{pmatrix} u_2 \\ u_1 \\ 0 \end{pmatrix} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_1 x_3 + \dot{x}_2 \\ \dot{x}_1 x_2 + \dot{x}_3 \end{pmatrix} \implies \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = \begin{pmatrix} u_2 \\ u_1 - u_2 x_3 \\ -u_2 x_2 \end{pmatrix}. \tag{6.3.9}$$

Let $x_1 = z \implies \dot{x}_1 = \dot{z} = u_2$. Introduce now following transformation in (6.3.9):

$$\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cosh x_1 & -\sinh x_1 \\ -\sinh x_1 & \cosh x_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$
$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cosh x_1 & \sinh x_1 \\ \sinh x_1 & \cosh x_1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}.$$

Now we have,

$$\begin{aligned} x &= x_2 \cosh x_1 + x_3 \sinh x_1, \\ \dot{x} &= \dot{x}_2 \cosh x_1 + x_2 \dot{x}_1 \sinh x_1 + \dot{x}_3 \sinh x_1 + x_3 \dot{x}_1 \cosh x_1, \\ &= (u_1 - u_2 x_3) \cosh x_1 + u_2 x_2 \sinh x_1 - u_2 x_2 \sinh x_1 + u_2 x_3 \cosh x_1 \\ &= u_1 \cosh x_1 \\ &= u_1 \cosh z. \end{aligned}$$

Similarly,

$$\begin{aligned} y &= x_2 \sinh x_1 + x_3 \cosh x_1, \\ \dot{y} &= \dot{x}_2 \sinh x_1 + x_2 \dot{x}_1 \cosh x_1 + \dot{x}_3 \cosh x_1 + x_3 \dot{x}_1 \sinh x_1, \\ &= (u_1 - u_2 x_3) \sinh x_1 + u_2 x_2 \cosh x_1 - u_2 x_2 \cosh x_1 + u_2 x_3 \sinh x_1 \\ &= u_1 \sinh x_1 \\ &= u_1 \sinh z. \end{aligned}$$

Hence,

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} u_1 \cosh z \\ u_1 \sinh z \\ u_2 \end{pmatrix}. \tag{6.3.10}$$

The local coordinates we calculated in (6.3.10) have nice physical interpretation. Our coordinates $(x(t), y(t))$ represent the coordinates of a hypothetical robot (unicycle) on the hyperbolic plane and $z(t)$ its orientation. Together we call it the configuration $q(x, y, z) \in \mathrm{SH}(2)$ of the dynamical system. Throughout the rest of the text we use M or $\mathrm{SH}(2)$ to denote the same configuration manifold of the dynamical system which is our Lie group as well. In that respect, the kinematic system on hyperbolic plane (6.3.5) and (6.3.10) has two velocities i.e., translational velocity and rotational velocity that we denote as u_1 and u_2. Clearly the

translational velocity along x and y direction on the hyperbolic plane are given as $u_1 \cosh z$ and $u_1 \sinh z$ respectively and rotational velocity is $\dot{z} = u_2$. Comparing it with (5.6.1) we have the following definition of the vector fields:

$$f_1(q) = \begin{pmatrix} \cosh z \\ \sinh z \\ 0 \end{pmatrix}, \qquad f_2(q) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{6.3.11}$$

Note that $f_1(q) = qA_2, \qquad f_2(q) = qA_1$ is the matrix representation of the same vector fields given in (6.3.11). Thus in coordinates $q(x, y, z)$ the basis vector fields are given as:

$$\begin{aligned} f_1(q) &= \begin{pmatrix} \cosh z & \sinh z & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}, \\ &= \cosh z \frac{\partial}{\partial x} + \sinh z \frac{\partial}{\partial y}, \end{aligned}$$

and

$$\begin{aligned} f_2(q) &= \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}, \\ &= \frac{\partial}{\partial z}. \end{aligned}$$

6.4 Pontryagin's Maximum Principle for Sub-Riemannian Problem on the Lie Group SH(2)

Consider dynamical system (5.6.1-5.6.4). By Cauchy Schwarz inequality proved in equation (4.3.4) we have:

$$(l(u))^2 = \left(\int_0^{t_1} \sqrt{u_1^2 + u_2^2} dt \right)^2 \leq t_1 \int_0^{t_1} (u_1^2 + u_2^2) dt.$$

Thus sub-Riemannian length functional minimization problem (5.6.3) is equivalent to the problem of minimizing the following energy functional with fixed t_1[15]:

$$J = \frac{1}{2} \int_0^{t_1} (u_1^2 + u_2^2) dt \to \min. \tag{6.4.1}$$

First step in optimal control problems is to obtain PMP form for (5.6.1),(5.6.2),(6.4.1) using coordinate free approach described in [20]. Consider control dependent Hamiltonian h_u^ν given as:

$$h_u^\nu(\lambda) = \langle \lambda, f_u(q) \rangle + \frac{\nu}{2}(u_1^2 + u_2^2), \quad q = \pi(\lambda), \quad \lambda \in T^*M. \tag{6.4.2}$$

where π is the canonical projection from cotangent bundle to the configuration manifold M. Now suppose Hamiltonians $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ correspond to basis vector fields f_i. Then (6.4.2) is given as:

$$h_u^\nu(\lambda) = u_1 h_1(\lambda) + u_2 h_2(\lambda) + \frac{\nu}{2}(u_1^2 + u_2^2), \qquad u \in \mathbb{R}^2. \tag{6.4.3}$$

Note that ν is a constant, $\lambda(t) \in T^*\mathrm{SH}(2)$ is the vector of costate variables and its dimension is same as that of the state vector. The costate variables play the role of a time varying Lagrange multiplier on the system dynamics acting as constraints on the optimal control system. We now state PMP for optimal control problem by using Theorem 12.3 [20] as:

Theorem 6.4.1. *Let $\tilde{u}(t)$ be optimal control and $\tilde{q}(t)$ be optimal trajectory for $t \in [0, t_1]$ and $h_u^\nu(\lambda)$ given by (6.4.3) be the Hamiltonian function for (5.6.1),(5.6.2),(6.4.1). Then, there exists a nontrivial pair:*

$$(\nu, \lambda_t) \neq 0, \qquad \nu \in \mathbb{R}, \quad \lambda_t \in T^*_{\tilde{q}(t)}M, \quad \pi(\lambda_t) = \tilde{q}(t),$$

where λ_t is a Lipschitzian curve and $\nu \in \{-1, 0\}$ is a number, for which following conditions hold for almost all time $t \in [0, t_1]$:

$$\dot{\lambda}_t = \overrightarrow{h}^\nu_{\tilde{u}(t)}(\lambda_t), \tag{6.4.4}$$

$$h^\nu_{\tilde{u}(t)}(\lambda_t) = \max_{u \in \mathbb{R}^2} h^\nu_{u(t)}(\lambda_t), \tag{6.4.5}$$

where $\overrightarrow{h}^\nu_{\tilde{u}(t)}(\lambda_t)$ is the Hamiltonian vector field corresponding to the maximized Hamiltonian function $h^\nu_{\tilde{u}(t)}$.

Abnormal Trajectories

The way we write Hamiltonian and obtain the necessary optimality conditions from PMP, there appear two kinds of trajectories i.e., abnormal and normal. Abnormal trajectories correspond to the case $\nu = 0$, meaning thereby that the trajectories are time optimal but do not minimize the sub-Riemannian length or the energy of the system. Abnormal trajectories that are not projections of normal ones are called strictly abnormal. Strictly abnormal trajectories point out to an important fact that there may be points on the manifold that cannot be joined by minimizing energy. The Hamiltonian (6.4.3) in this case can be written as:

$$h_u^0(\lambda) = u_1 h_1(\lambda) + u_2 h_2(\lambda). \tag{6.4.6}$$

Theorem 6.4.2. *All abnormal extremal trajectories for problem* (5.6.1),(5.6.2),(6.4.1) *are constant.*

Proof. Let λ_t be the abnormal extremal. Applying the PMP maximization condition on equation (6.4.6):

$$\nabla h_u^0(\lambda) = \begin{pmatrix} h_1(\lambda) \\ h_2(\lambda) \end{pmatrix} = 0, \tag{6.4.7}$$

$$\implies h_1(\lambda_t) = h_2(\lambda_t) \equiv 0. \tag{6.4.8}$$

Differentiating equation (6.4.8) w.r.t. Hamiltonian vector field :

$$\begin{aligned} \dot{h}_1 &= \left\{h_u^0, h_1\right\} = \{u_1 h_1 + u_2 h_2, h_1\} = u_1 \{h_1, h_1\} + u_2 \{h_2, h_1\} = u_2 h_0, \\ \dot{h}_2 &= \left\{h_u^0, h_2\right\} = \{u_1 h_1 + u_2 h_2, h_2\} = u_1 \{h_1, h_2\} + u_2 \{h_2, h_2\} = -u_1 h_0. \end{aligned}$$

Note that Poisson bracket and Lie bracket of sh(2) follow the same multiplication rule. Since $h_1 = h_2 = 0 \implies \dot{h}_1 = \dot{h}_2 = 0$, therefore,

$$\begin{aligned} u_2(t) h_0(\lambda_t) &= u_1(t) h_0(\lambda_t) \equiv 0, \\ \implies u_1^2(t) h_0^2 + u_2^2(t) h_0^2 &= 0. \end{aligned}$$

If $h_0(\lambda_t) = 0$ for some λ_t, then $h_1(\lambda_t) = h_2(\lambda_t) = h_0(\lambda_t) \equiv 0$ which means $\lambda_t = 0$. Since we are considering the case for $\nu = 0$, this is impossible. Therefore $u_1^2(t) + u_2^2(t) = 0 \implies u_1 = u_2 = 0$ and hence the abnormal extremal trajectories are constant. □

Normal Trajectories

When $\nu = -1$ (6.4.3) the trajectories given by PMP are called normal extremal trajectories. The Hamiltonian in this case is given as:

$$H = h_u^{-1}(\lambda) = u_1 h_1(\lambda) + u_2 h_2(\lambda) - \frac{1}{2}\left(u_1^2 + u_2^2\right), \qquad u \in \mathbb{R}^2. \tag{6.4.9}$$

Applying the first order optimality conditions w.r.t the controls:

$$\frac{\partial H}{\partial u} = \begin{pmatrix} h_1 - u_1 \\ h_2 - u_2 \end{pmatrix} = 0, \tag{6.4.10}$$

$$\implies u_1 = h_1, \quad u_2 = h_2. \tag{6.4.11}$$

For $u_i = 0$, the normal extremal trajectories are constant which in turn concludes that the abnormal trajectories are not strictly abnormal. The maximized Hamiltonian corresponding to Hamiltonian system $\dot{\lambda} = \overrightarrow{H}(\lambda)$, $\lambda \in T^*M$ for normal case is $H = \frac{1}{2}\left(h_1^2 + h_2^2\right) \geq 0$. As can be seen, $H > 0$ in case of non-constant normal extremals. The maximized Hamiltonian function has another interesting property in normal case i.e., it is homogeneous w.r.t. h_1, h_2 and therefore we can consider its trajectories for the level surface $H = \frac{1}{2}$. In this case, the initial covector λ is contained in the phase cylinder C and is given as:

$$C = T_{q_0}^* M \cap \left\{H(\lambda) = \frac{1}{2}\right\} = \left\{(h_1, h_2, h_0) \in \mathbb{R}^3 \quad | \quad h_1^2 + h_2^2 = 1\right\}. \tag{6.4.12}$$

Differentiating (6.4.11) w.r.t. Hamiltonian vector field we get:

$$\begin{aligned}
\dot{h}_1 &= \{H, h_1\} = \left\{\frac{1}{2}\left(h_1^2 + h_2^2\right), h_1\right\} = h_2\{h_2, h_1\} = h_2 h_0, \\
\dot{h}_2 &= \{H, h_2\} = \left\{\frac{1}{2}\left(h_1^2 + h_2^2\right), h_2\right\} = h_1\{h_1, h_2\} = -h_1 h_0, \\
\dot{h}_0 &= \{H, h_0\} = \left\{\frac{1}{2}\left(h_1^2 + h_2^2\right), h_0\right\} = h_1\{h_1, h_0\} + h_2\{h_2, h_0\} = h_1 h_2.
\end{aligned}$$

Hence, complete Hamiltonian system in normal case is given as:

$$\begin{pmatrix} \dot{h}_1 \\ \dot{h}_2 \\ \dot{h}_0 \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} h_2 h_0 \\ -h_1 h_0 \\ h_1 h_2 \\ h_1 \cosh z \\ h_1 \sinh z \\ h_2 \end{pmatrix}. \tag{6.4.13}$$

6.5 Vertical Subsystem

The entire machinery of PMP rests on cotangent lift of the problem. The differential equations are lifted to the cotangent bundle and the resulting system is called the vertical subsystem. The vertical subsystem is solved and through appropriate exponential mapping the original system of ODEs called the horizontal subsystem is solved on configuration manifold. Here we state and prove an important theorem regarding the vertical subsystem. The vertical subsystem in appropriate coordinates represents a mathematical pendulum.

Theorem 6.5.1. *Vertical subsystem of the Hamiltonian system (6.4.13) in normal case is a mathematical pendulum.*

Proof. Introduce following coordinates transformation:

$$h_1 = \cos\alpha, \quad h_2 = \sin\alpha. \tag{6.5.1}$$

Thus,

$$\begin{aligned}
\dot{h}_1 &= -\sin\alpha\dot{\alpha} = \sin\alpha.h_0, \\
\dot{\alpha} &= -h_0.
\end{aligned} \tag{6.5.2}$$

Similarly,

$$\dot{h}_0 = \cos\alpha\sin\alpha = \frac{1}{2}\sin 2\alpha. \tag{6.5.3}$$

Let us introduce another change of coordinates:

$$\gamma = 2\alpha \in 2S^1 = \mathbb{R}/4\pi\mathbb{Z}, \quad c = -2h_0 \in \mathbb{R}, \tag{6.5.4}$$

$$\implies \dot{\gamma} = 2\dot{\alpha} = -2h_0 = c,$$

and

$$\dot{c} = -2\dot{h}_0 = -2h_1h_2 = -2\cos\alpha\sin\alpha = -\sin 2\alpha = -\sin\gamma.$$

Thus,

$$\begin{pmatrix} \dot{\gamma} \\ \dot{c} \end{pmatrix} = \begin{pmatrix} c \\ -\sin\gamma \end{pmatrix}. \tag{6.5.5}$$

It can be easily seen that (6.5.5) represents a mathematical pendulum. □

Note that vertical subsystem being a pendulum affords integration of vertical subsystem and hence that of the complete Hamiltonian system in terms of Jacobi elliptic functions defined in Appendix 1. In fact, standard texts on elliptic functions such as [71] introduce Jacobi elliptic functions via the example of a pendulum.

6.6 Integrability

Theorem 6.6.1. *The normal Hamiltonian system (6.4.13) is completely integrable.*

Proof. From the theory of integrable systems we know that a $2n$-dimensional Hamiltonian system is called completely integrable if there are n Poisson-commuting smooth integrals $f_1, \ldots, f_n$ such that:

$$\{f_i, f_j\} = 0, \qquad i,\, j = 1, \ldots, n, \tag{6.6.1}$$

whose differentials are independent in an open dense subset U of phase space T^*M i.e., $U \subseteq T^*M$. In order to prove that (6.4.13) is integrable, we follow the approach given in [72]. The Lie Algebra $L = sh(2) = span\,\{A_1, A_2, A_3\}$. The left invariant vector fields $X_i = f_i$ read as:

$$X_1 = \begin{pmatrix} \cosh z \\ \sinh z \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & \cosh z \\ 0 & 0 & \sinh z \\ 0 & 0 & 0 \end{pmatrix} = q.A_2, \tag{6.6.2}$$

$$X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} \sinh z & \cosh z & 0 \\ \cosh z & \sinh z & 0 \\ 0 & 0 & 0 \end{pmatrix} = q.A_1, \tag{6.6.3}$$

$$X_3 = \begin{pmatrix} \sinh z \\ \cosh z \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & \sinh z \\ 0 & 0 & \cosh z \\ 0 & 0 & 0 \end{pmatrix} = q.A_3. \tag{6.6.4}$$

The right invariant vector fields Y_i read as [72]:

$$Y_1 = A_2 q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.6.5}$$

$$Y_2 = A_1 q = \begin{pmatrix} \sinh z & \cosh z & y \\ \cosh z & \sinh z & x \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.6.6}$$

$$Y_3 = A_3 q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.6.7}$$

In coordinates x,y,z the left invariant vector fields read as:

$$X_1 = \partial_x \cosh z + \partial_y \sinh z, \tag{6.6.8}$$

$$X_2 = \partial_z, \tag{6.6.9}$$

$$X_3 = \partial_x \sinh z + \partial_y \cosh z. \tag{6.6.10}$$

Now using (6.6.8)–(6.6.10)we have:

$$\partial_x = X_1 \cosh z - X_3 \sinh z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.6.11}$$

$$\partial_y = -X_1 \sinh z + X_3 \cosh z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.6.12}$$

$$\partial_z = X_2 = \begin{pmatrix} \sinh z & \cosh z & 0 \\ \cosh z & \sinh z & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.6.13}$$

The left invariant Hamiltonians $h_i(\lambda) = \langle \lambda, X_i \rangle$ in canonical coordinates $(x, y, z, \psi_x, \psi_y, \psi_z)$ are given as:

$$h_1 = \psi_x \cosh z + \psi_y \sinh z, \tag{6.6.14}$$

$$h_2 = \psi_z, \tag{6.6.15}$$

$$h_0 = \psi_x \sinh z + \psi_y \cosh z. \tag{6.6.16}$$

Similarly the right invariant Hamiltonians $g_i(\lambda) = \langle \lambda, Y_i \rangle$ are written as:

$$g_1 = \psi_x, \tag{6.6.17}$$

$$g_2 = \psi_x y + \psi_y x + \psi_z, \tag{6.6.18}$$

$$g_3 = \psi_y. \tag{6.6.19}$$

It can be verified that left invariant Hamiltonian $H = \frac{1}{2}(h_1^2 + h_2^2)$ Poisson commutes with the right-invariant Hamiltonians. Thus Hamiltonian system (6.4.13) has algebra of integrals:

$$I = \text{span}_g = (H, g_1, g_2, g_3)\,. \tag{6.6.20}$$

The non-zero Poisson bracket in the basis elements of I are:

$$\{g_1, g_2\} = -g_2 \qquad \{g_2, g_3\} = g_1. \tag{6.6.21}$$

The integrals H, g_1, g_3 Poisson commute and hence are in involution. It only remains to show that the integrals H, g_1, g_3 are functionally independent on some dense subset U of the phase space T^*M i.e., $U \subseteq T^*M$. Consider the Jacobian J of the integrals of motion:

$$\mathrm{J} = \begin{pmatrix} \nabla H \\ \nabla g_1 \\ \nabla g_3 \end{pmatrix},$$

$$\mathrm{J} = \begin{pmatrix} 0 & 0 & l & m & n & \psi_z \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{6.6.22}$$

where $l = \sinh z \cosh z \left(\psi_x^2 + \psi_y^2\right) + \psi_x \psi_y \left(\cosh^2 z + \sinh^2 z\right)$, $m = \psi_x \cosh^2 z + \psi_y \cosh z \sinh z$, $n = \psi_y \sinh^2 z + \psi_x \cosh z \sinh z$. In terms of left invariant Hamiltonians h_i (6.6.22) reads as follows:

$$\mathrm{J} = \begin{pmatrix} 0 & 0 & h_1 h_0 & h_1 \cosh z & h_1 \sinh z & \psi_z \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{6.6.23}$$

It can be easily seen that in J (6.6.23) is full rank i.e., H, g_1, g_3 are functionally independent on subset $U = T^*M/S$ where $S = \{\lambda \in T^*M \mid h_2 = h_0 = 0\}$. Restriction of Hamiltonian system (6.4.13) to invariant manifold S becomes:

$$\dot{h}_1 = \dot{h}_2 = \dot{h}_3 = 0, \qquad \dot{x} = h_1 \cosh z, \qquad \dot{y} = h_1 \sinh z.$$

Hence for 6-dimensional Hamiltonian system (6.4.13) the system is completely integrable on T^*M. Note that J also loses rank on $S = \{\lambda \in T^*M \mid h_1 = h_2 = 0\}$ but in that case (6.4.13) becomes identically zero and we do not consider that trivial case. Hence the research objective 1 of the sub-Riemannian problem on Lie group SH(2) is achieved. □

6.7 Transformation of Lie Group SH(2) to SOLV$^-$

Claim 6.7.1. Sub-Riemannian problem was investigated by A. D. Mazhitova in [52] on Lie group SOLV$^-$. We prove that sub-Riemannian problem on SH(2) can be transformed into equivalent problem on SOLV$^-$.

Proof. The matrix representation for Lie Group SOLV^- is given as:

$$\mathrm{SOLV}^- = \left\{ \begin{pmatrix} e^{-z} & 0 & x \\ 0 & e^{z} & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}. \tag{6.7.1}$$

The corresponding representation for SH(2) is given as:

$$M = \mathrm{SH}(2) = \left\{ \begin{pmatrix} \cosh z & \sinh z & x \\ \sinh z & \cosh z & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}. \tag{6.7.2}$$

Consider a following orthogonal transformation T:

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}. \tag{6.7.3}$$

Applying transformation to equation (6.7.2):

$$\begin{aligned} TMT^T &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} \cosh z & \sinh z & x \\ \sinh z & \cosh z & y \\ 0 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}^T, \\ &= \begin{pmatrix} e^{-z} & 0 & \frac{1}{2}(x-y) \\ 0 & e^{z} & \frac{1}{2}(x+y) \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{6.7.4}$$

Now let $\frac{1}{2}(x-y) = w$ and $\frac{1}{2}(x+y) = v$ in equation (6.7.4) and the resulting matrix becomes:

$$TMT^T = \begin{pmatrix} e^{-z} & 0 & w \\ 0 & e^{z} & v \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.7.5}$$

Note that 6.7.5 is same as the matrix representation for SOLV^-. Same orthogonal transformation applies to Lie algebra $\mathrm{solv}^- \to \mathrm{sh}(2)$. Hence proved. □

6.8 Chapter Summary

We were able to establish that the sub-Riemannian problem on SH(2) is completely controllable. The notion of integrability is important as it ensures that system can be integrated analytically and some closed form solution will appear. The closed form solutions allow further optimality analysis which is otherwise intractable. Once integrability is established, we can attempt the analytical integration of the Hamiltonian system which shall complete the parametrization of extremal trajectories.

Chapter 7

Parametrization of the Extremal Trajectories

This chapter pertains to one of the main results of the sub-Riemannian problem on group of motions of hyperbolic plane. In previous chapter we proved that the vertical subsystem of the Hamiltonian system is a mathematical pendulum. Based on this fact we define suitable elliptic coordinates in which we integrate the Hamiltonian system and parametrize the extremal trajectories. The extremal trajectories are parametrized in terms of Jacobi elliptic functions. Qualitative analysis reveals that extremal trajectories have inflection points and cusps.

7.1 Decomposition of the Initial Phase Cylinder

Hamiltonian system for normal trajectories was given in (6.4.13). Under the transformations introduced in (6.5.1),(6.5.4), the horizontal subsystem can be written as:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} h_1 \cosh z \\ h_1 \sinh z \\ h_2 \end{pmatrix} = \begin{pmatrix} \cos\frac{\gamma}{2} \cosh z \\ \cos\frac{\gamma}{2} \sinh z \\ \sin\frac{\gamma}{2} \end{pmatrix}. \tag{7.1.1}$$

Vertical subsystem being a pendulum, the initial phase cylinder C of vertical subsystem can be decomposed on the basis of the energy of the pendulum that in turn corresponds to its various trajectories. The decomposition hence leads to fairly standard subsets of C [1]. The decomposition procedure and corresponding subsets are as follows.
The total energy integral of the pendulum obtained in (6.5.5) is given as:

$$E = \frac{c^2}{2} - \cos\gamma = 2h_0^2 - h_1^2 + h_2^2, \qquad E \in [-1, +\infty). \tag{7.1.2}$$

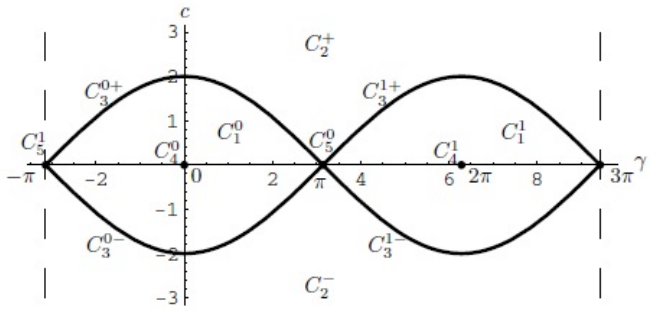

Figure 7.1: Decomposition of the Phase Cylinder and the Connected Subsets [1]

The initial phase cylinder (6.4.12) may be decomposed into following subsets based upon the pendulum energy that correspond to various pendulum trajectories:

$$C = \bigcup_{i=1}^{5} C_i,$$

where

$$C_1 = \{\lambda \in C | E \in (-1, 1)\}, \tag{7.1.3}$$
$$C_2 = \{\lambda \in C | E \in (1, \infty)\}, \tag{7.1.4}$$
$$C_3 = \{\lambda \in C | E = 1, c \neq 0\}, \tag{7.1.5}$$
$$C_4 = \{\lambda \in C | E = -1\} = \{(\gamma, c) \in C | \gamma = 2\pi n, c = 0\}, \quad n \in \mathbb{N}, \tag{7.1.6}$$
$$C_5 = \{\lambda \in C | E = 1\} = \{(\gamma, c) \in C | \gamma = 2\pi n + \pi, c = 0\}, \quad n \in \mathbb{N}. \tag{7.1.7}$$

Continuing the approach taken in [1] the subsets C_i may be further decomposed as:

$$\begin{aligned}
C_1 &= \cup_{i=0}^{1} C_1^i, \quad C_1^i = \{(\gamma, c) \in C_1 | \operatorname{sgn}(\cos(\gamma/2)) = (-1)^i\}, \\
C_2 &= C_2^+ \cup C_2^-, \quad C_2^{\pm} = \{(\gamma, c) \in C_2 | \operatorname{sgn} c = \pm 1\}, \\
C_3 &= \cup_{i=0}^{1} (C_3^{i+} \cup C_3^{i-}), \quad C_3^{i\pm} = \{(\gamma, c) \in C_3 | \operatorname{sgn}(\cos(\gamma/2)) = (-1)^i, \operatorname{sgn} c = \pm 1\}, \\
C_4 &= \cup_{i=0}^{1} C_4^i, \quad C_4^i = \{(\gamma, c) \in C | \gamma = 2\pi i, c = 0\}, \\
C_5 &= \cup_{i=0}^{1} C_5^i, \quad C_5^i = \{(\gamma, c) \in C | \gamma = 2\pi i + \pi, c = 0\}.
\end{aligned}$$

In all of the above $i = 0, 1$. The phase portrait of the pendulum and corresponding decomposition of initial phase cylinder C is depicted in Figure 7.1. Note that such decomposition is important for parametrization of extremal trajectories in terms of Jacobi elliptic functions.

7.2 Rectification

We introduce here a subsidiary concept of rectification related to dynamical systems that shall prove vital in the integration of the Hamiltonian system equation (6.4.13). Intuitively,

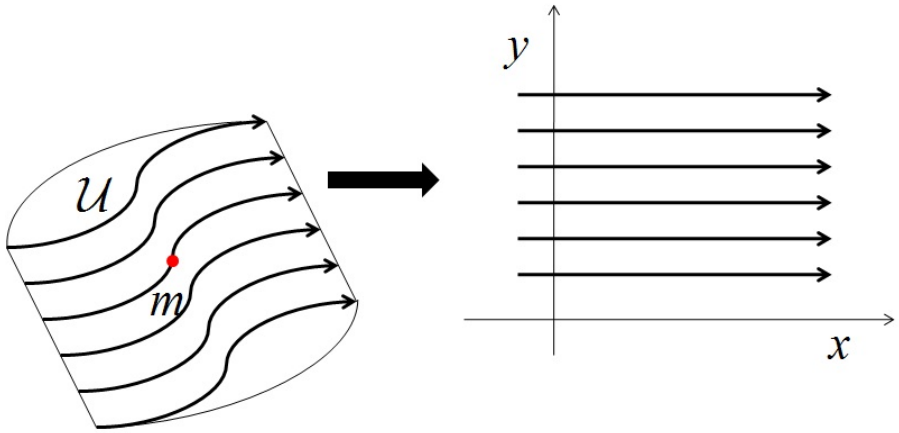

Figure 7.2: Rectification of Flow of Dynamical System

rectification means that away from singular points the flow lines of a dynamical system become parallel or dynamics of a point $m \in M$ in its small neighborhood is a straight line. Depending upon the dynamics of the system the small patches can glued together to form a bigger patch. Occasionally it is possible to glue enough small patches to cover the entire phase space M meaning thereby that the dynamical system is integrable. The concept is based on the rectification theorem which states that:
Let $m \in M$ be a regular point of a C^r flow $f^t | r \geq 0$ on a 2D manifold M then there exists a neighborhood $\mathcal{U}$ of m and a C^r diffeomorphism $\mathcal{U} \to \mathbb{R}^2$ carrying the arcs in $\mathcal{U}$ of trajectories into trajectories of the dynamical system $\dot{x} = 1, \qquad \dot{y} = 0$ on $\mathbb{R}^2$ with the preservation of direction with time.

7.3 Elliptic Coordinates

We reparametrize the time and energy of the pendulum in terms of of elliptic coordinates (φ, k) respectively on the domain $\cup_{i=1}^{3} C_i \subset C$ as given in [1], [53]. This allows us to introduce Jacobi elliptic functions $\mathrm{sn}(\varphi, k)$, $\mathrm{cn}(\varphi, k)$, $\mathrm{dn}(\varphi, k)$, $\mathrm{am}(\varphi, k)$, and $\mathrm{E}(\varphi, k) = \int_0^{\varphi} \mathrm{dn}^2(t, k) dt$ while integrating the extremals and extremal trajectories corresponding to vertical and horizontal subsystem respectively. Jacobi elliptic functions are described in Appendix 1. Detailed description may be found in [71].

Case 1 : $\lambda = (\varphi, k) \in C_1$

$$k = \sqrt{\frac{E+1}{2}} = \sqrt{\sin^2\frac{\gamma}{2} + \frac{c^2}{4}} \in (0,1), \tag{7.3.1}$$

$$\sin\frac{\gamma}{2} = s_1 k\,\mathrm{sn}(\varphi,k), \quad s_1 = \mathrm{sgn}\left(\cos\frac{\gamma}{2}\right), \tag{7.3.2}$$

$$\cos\frac{\gamma}{2} = s_1 \mathrm{dn}(\varphi,k), \tag{7.3.3}$$

$$\frac{c}{2} = k\,\mathrm{cn}(\varphi,k), \quad \varphi \in [0, 4K(k)]. \tag{7.3.4}$$

Proposition 7.3.1. *The elliptic coordinates cause the flow of vertical subsystem to be rectified i.e., the flow lines become parallel to each other.*

Proof. Using (7.3.1),

$$k^2 = \sin^2\frac{\gamma}{2} + \frac{c^2}{4}. \tag{7.3.5}$$

Taking the time derivative of (7.3.5),

$$2k\dot{k} = 2\sin\frac{\gamma}{2}\cos\frac{\gamma}{2}\frac{\dot{\gamma}}{2} + \frac{c\dot{c}}{2}. \tag{7.3.6}$$

Using (6.5.5),

$$2k\dot{k} = \frac{\dot{\gamma}}{2}\sin\gamma - \frac{\dot{\gamma}}{2}\sin\gamma = 0.$$

In previous equation, either $k = 0$ or $\dot{k} = 0$. Since $k \in (0,1)$, therefore $k = 0$ is impossible and therefore:

$$\dot{k} = 0. \tag{7.3.7}$$

Using (7.3.4) and the derivatives of elliptic functions defined in [71],

$$\begin{aligned}
\frac{d}{dt}\left(\frac{c}{2}\right) &= \frac{d}{dt}k\mathrm{cn}(\varphi,k),\\
\frac{\dot{c}}{2} &= k\frac{d}{d\varphi}\mathrm{cn}(\varphi,k).\frac{d\varphi}{dt} + k\frac{d}{dk}\mathrm{cn}(\varphi,k).\frac{dk}{dt} + \mathrm{cn}(\varphi,k).\frac{dk}{dt},\\
-\sin\gamma &= -2k\mathrm{sn}(\varphi,k)\mathrm{dn}(\varphi,k)\dot{\varphi}.
\end{aligned}$$

because $\frac{dk}{dt} = 0$. Now,

$$\dot{\varphi} = \frac{\sin\gamma}{2k\,\mathrm{sn}(\varphi,k).\mathrm{dn}(\varphi,k)}. \tag{7.3.8}$$

Now using (7.3.2),(7.3.3):

$$\begin{aligned}
\sin\frac{\gamma}{2}\cos\frac{\gamma}{2} &= s_1 k\,\mathrm{sn}(\varphi,k).s_1\mathrm{dn}(\varphi,k),\\
2\sin\frac{\gamma}{2}\cos\frac{\gamma}{2} &= 2s_1^2 k\,\mathrm{sn}(\varphi,k).\mathrm{dn}(\varphi,k),\\
\sin\gamma &= 2k\,\mathrm{sn}(\varphi,k).\mathrm{dn}(\varphi,k).
\end{aligned}$$

Thus (7.3.8) becomes:

$$\dot{\varphi} = 1. \tag{7.3.9}$$

□

Case 2 : $\lambda = (\varphi, k) \in C_2$

$$k = \sqrt{\frac{2}{E+1}} = \sqrt{\frac{1}{\sin^2\frac{\gamma}{2} + \frac{c^2}{4}}} \in (0,1), \tag{7.3.10}$$

$$\sin\frac{\gamma}{2} = s_2 \mathrm{sn}\left(\frac{\varphi}{k}, k\right), \quad s_2 = \mathrm{sgn}(c), \tag{7.3.11}$$

$$\cos\frac{\gamma}{2} = \mathrm{cn}\left(\frac{\varphi}{k}, k\right), \tag{7.3.12}$$

$$\frac{c}{2} = \frac{s_2}{k}\mathrm{dn}\left(\frac{\varphi}{k}, k\right), \quad \varphi \in [0, 4kK(k)]\,. \tag{7.3.13}$$

Case 3 : $\lambda = (\varphi, k) \in C_3$

$$k = 1, \tag{7.3.14}$$

$$\sin\frac{\gamma}{2} = s_1 s_2 \tanh\varphi, \quad s_1 = \mathrm{sgn}\left(\cos\frac{\gamma}{2}\right), \quad s_2 = \mathrm{sgn}(c), \tag{7.3.15}$$

$$\cos\frac{\gamma}{2} = s_1/\cosh\varphi, \tag{7.3.16}$$

$$\frac{c}{2} = s_2/\cosh\varphi, \quad \varphi \in (-\infty, \infty). \tag{7.3.17}$$

Rectification of flow for cases 2 and 3 can be proved using the same procedure as for case 1. Note that rectification has tremendously simplified the integration of the vertical subsystem. Since there is a diffeomorphism between the vertical subsystem in original coordinates and the elliptic coordinates, hence the elliptic coordinates represent the same system in a simpler way.

7.4 Integration of Vertical Subsystem

Due to rectification of flow of vertical subsystem in elliptic coordinates, the integration of vertical subsystem becomes trivial i.e., $\varphi_t = t + \varphi$ and $k = constant$. Here φ is the initial value of φ_t at $t = 0$. Note that in the definition of Jacobi elliptic functions, the variable φ_t will be used as φ is merely a constant of integration from here onward. Subscript t is used to explicitly signify the time dependent variables.

7.5 Integration of Horizontal Subsystem

We now consider integration of the horizontal subsystem (7.1.1) for cases 1-3 noted above. Assuming zero initial state i.e., $x(0) = y(0) = z(0) = 0$.

Case 1 : $\lambda = (\varphi, k) \in C_1$

Theorem 7.5.1. *In case 1 extremal trajectories are parametrized as follows:*

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \frac{s_1}{2}\left[\left(w + \frac{1}{w(1-k^2)}\right)[\mathrm{E}(\varphi_t) - \mathrm{E}(\varphi)] + \left(\frac{k}{w(1-k^2)} - kw\right)[\mathrm{sn}\,\varphi_t - \mathrm{sn}\,\varphi]\right] \\ \frac{1}{2}\left[\left(w - \frac{1}{w(1-k^2)}\right)[\mathrm{E}(\varphi_t) - \mathrm{E}(\varphi)] - \left(\frac{k}{w(1-k^2)} + kw\right)[\mathrm{sn}\,\varphi_t - \mathrm{sn}\,\varphi]\right] \\ s_1 \ln\left[(\mathrm{dn}\,\varphi_t - k\mathrm{cn}\,\varphi_t).w\right] \end{pmatrix}, \tag{7.5.1}$$

where $w = \frac{1}{\mathrm{dn}\varphi - k\mathrm{cn}\varphi}$.

Proof. From (7.1.1) consider $\dot{z} = \sin\frac{\gamma}{2} = s_1 k\,\mathrm{sn}(\varphi, k)$. The solution to this ODE can be written as:

$$\frac{dz}{dt} = \frac{dz}{d\varphi_t}.\frac{d\varphi_t}{dt} = \frac{dz}{d\varphi_t}.\dot{\varphi} = \frac{dz}{d\varphi_t} = s_1 k\,\mathrm{sn}(\varphi_t, k), \tag{7.5.2}$$

$$\int_0^{z_t} dz = \int_{\varphi}^{\varphi_t} s_1 k\,\mathrm{sn}\varphi_t\, d\varphi_t,$$

where u is the variable of integration. Using [73], equation (7.5.2) becomes:

$$z_t = s_1 \ln(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t) - s_1 \ln(\mathrm{dn}\varphi - k\mathrm{cn}\varphi). \tag{7.5.3}$$

Let $\ln w = -\ln(\mathrm{dn}\varphi - k\mathrm{cn}\varphi)$, $w = \frac{1}{\mathrm{dn}\varphi - k\mathrm{cn}\varphi}$. Then (7.5.3) becomes,

$$z_t \;=\; s_1 \ln[(\mathrm{dn}\varphi_t - k\mathrm{cn}\,\varphi_t).w]. \tag{7.5.4}$$

From (7.1.1) now consider,

$$\dot{x} \;=\; \cos\frac{\gamma}{2}\cosh z = s_1 \mathrm{dn}\varphi_t \cosh\left(s_1 \ln\left[(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t).w\right]\right), \tag{7.5.5}$$

$$\dot{x} \;=\; \frac{s_1}{2}\mathrm{dn}\varphi_t \left(e^{\ln[(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w]} + e^{-\ln[(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w]}\right),$$

$$\dot{x} \;=\; \frac{s_1}{2}\mathrm{dn}\varphi_t \left((\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w + \left[(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w\right]^{-1}\right),$$

$$\dot{x} \;=\; \frac{s_1}{2}\mathrm{dn}\varphi_t \left((\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w + \frac{1}{\left[(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t)^{s_1}.w\right]}\right). \tag{7.5.6}$$

Thus (7.5.6) becomes,

$$dx = \frac{s_1}{2}\left(w.\mathrm{dn}^2\varphi_t - kw.\mathrm{dn}\varphi_t\mathrm{cn}\varphi_t + \frac{\mathrm{dn}\varphi_t}{(\mathrm{dn}\varphi_t - k\mathrm{cn}\varphi_t).w}\right) d\varphi_t. \tag{7.5.7}$$

Now using the following identities of Elliptic functions:

$$\mathrm{sn}'\varphi = \mathrm{cn}\varphi\,\mathrm{dn}\varphi, \tag{7.5.8}$$

$$\int_0^{\varphi} \mathrm{dn}^2 u\, du = E(\varphi), \tag{7.5.9}$$

where $E(\varphi)$ is the elliptic integral of second type.

$$\mathrm{sn}^2\varphi + \mathrm{cn}^2\varphi = 1, \tag{7.5.10}$$

and

$$\mathrm{dn}^2\varphi + k^2\mathrm{sn}^2\varphi = 1, \tag{7.5.11}$$

we can write

$$\mathrm{dn}^2\varphi - k^2\mathrm{cn}^2\varphi = \mathrm{dn}^2\varphi - k^2(1-\mathrm{sn}^2\varphi) = \mathrm{dn}^2\varphi - k^2 + k^2\mathrm{sn}^2\varphi = 1 - k^2. \tag{7.5.12}$$

Thus (7.5.7) can be integrated as:

$$x_t = \frac{s_1}{2}\left[w\int_{\varphi}^{\varphi_t}\mathrm{dn}^2\varphi_t d\varphi_t - kw\int_{\varphi_0}^{\varphi}\mathrm{dn}\,\varphi_t\mathrm{cn}\,\varphi_t d\varphi_t + \frac{1}{w}\int_{\varphi_0}^{\varphi}\frac{\mathrm{dn}^2\varphi_t + k\mathrm{cn}\,\varphi_t\mathrm{dn}\,\varphi_t}{\mathrm{dn}^2\varphi_t - k^2\mathrm{cn}^2\varphi_t}d\varphi_t\right]. \tag{7.5.13}$$

Now using the standard identities of elliptic functions result of integration of (7.5.13) can be written as:

$$x_t = \frac{s_1}{2}\left[\left(w + \frac{1}{w\,(1-k^2)}\right)\left[\mathrm{E}(\varphi_t) - \mathrm{E}(\varphi)\right] + \left(\frac{k}{w\,(1-k^2)} - kw\right)\left[\mathrm{sn}\,\varphi_t - \mathrm{sn}\,\varphi\right]\right]. \tag{7.5.14}$$

From (7.1.1) now consider,

$$\begin{aligned}
\dot{y} &= \cos\frac{\gamma}{2}\sinh z_t = s_1\mathrm{dn}\,\varphi\sinh\left(s_1\ln\left[(\mathrm{dn}\,\varphi - k\mathrm{cn}\,\varphi).w\right]\right),\\
\dot{y} &= s_1^2\mathrm{dn}\,\varphi\sinh\left(s_1\ln\left[(\mathrm{dn}\,\varphi - k\mathrm{cn}\,\varphi).w\right]\right),\\
\dot{y} &= \mathrm{dn}\,\varphi\sinh\left(s_1\ln\left[(\mathrm{dn}\,\varphi - k\mathrm{cn}\,\varphi).w\right]\right).
\end{aligned} \tag{7.5.15}$$

The integration follows the same pattern as described above and hence final result of integration of (7.5.15) can be written as:

$$y_t = \frac{1}{2}\left[\left(w - \frac{1}{w\,(1-k^2)}\right)\left[\mathrm{E}(\varphi_t) - \mathrm{E}(\varphi)\right] - \left(\frac{k}{w\,(1-k^2)} + kw\right)\left[\mathrm{sn}\,\varphi_t - \mathrm{sn}\,\varphi\right]\right]. \tag{7.5.16}$$

□

Case 2 : $\lambda = (\varphi, k) \in C_2$

Theorem 7.5.2. *Consider horizontal system (7.1.1) for case 2 (7.3.10)–(7.3.13) and substitute $\psi = \frac{\varphi}{k}$ and $\psi = \frac{\varphi}{k} = \psi + \frac{t}{k}$, the integration results can be summarized as:*

$$\begin{aligned}
x_t &= \frac{1}{2}\left(\frac{1}{w(1-k^2)} - w\right)\left[\mathrm{E}(\psi_t) - \mathrm{E}(\psi) - k'^2\,(\psi_t - \psi)\right]\\
&+ \frac{1}{2}\left(kw + \frac{k}{w(1-k^2)}\right)\left[\mathrm{sn}\,\psi_t - \mathrm{sn}\,\psi\right],\\
y_t &= -\frac{s_2}{2}\left(\frac{1}{w(1-k^2)} + w\right)\left[\mathrm{E}(\psi_t) - \mathrm{E}(\psi) - k'^2(\psi_t - \psi)\right]\\
&+ \frac{s_2}{2}\left(kw - \frac{k}{w(1-k^2)}\right)\left[\mathrm{sn}\,\psi_t - \mathrm{sn}\,\psi\right],\\
z_t &= s_2\ln\left[(\mathrm{dn}\,\psi_t - k\mathrm{cn}\,\psi_t)\,.w\right],
\end{aligned} \tag{7.5.17}$$

where $w = \frac{1}{\mathrm{dn}\,\psi - k\mathrm{cn}\,\psi}$.

Proof. Proof follows from the procedure outlined in case 1. □

Case 3 : $\lambda = (\varphi, k) \in C_3$

Theorem 7.5.3. *In this case,* $k = 1$. *Integration results are summarized as:*

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \frac{s_1}{2}\left[\frac{1}{w}(\varphi_t - \varphi) + w(\tanh\varphi_t - \tanh\varphi)\right] \\ \frac{s_2}{2}\left[\frac{1}{w}(\varphi_t - \varphi) - w(\tanh\varphi_t - \tanh\varphi)\right] \\ -s_1 s_2 \ln[w\, sech\, \varphi_t] \end{pmatrix}, \tag{7.5.18}$$

and $w = \cosh\varphi$.

Proof. Consider horizontal system (7.1.1) for case 3 (7.3.14)–(7.3.17):

$$\begin{aligned} \dot{z} &= \sin\frac{\gamma}{2} = s_1 s_2 \tanh\varphi_t, \\ z_t &= -s_1 s_2[\ln(\operatorname{sech}\varphi_t) - \ln(\operatorname{sech}\varphi)]. \end{aligned}$$

Let $-\ln(\operatorname{sech}\varphi) = \ln w$, $w = \cosh\varphi$, then:

$$z_t = -s_1 s_2 \ln[w \operatorname{sech}\varphi_t]. \tag{7.5.19}$$

From (7.1.1) now consider,

$$\begin{aligned} \dot{x} &= \cos\frac{\gamma}{2}\cosh z_t = s_1 \operatorname{sech}\varphi_t \cosh\left(-s_1 s_2 \ln[w\operatorname{sech}\varphi_t]\right), \\ \dot{x} &= \frac{s_1 \operatorname{sech}\varphi_t}{2}\left[e^{\ln[w\operatorname{sech}\varphi_t]} + e^{-\ln[w\operatorname{sech}\varphi_t]}\right], \\ dx &= \frac{s_1 \operatorname{sech}\varphi_t}{2}\left[\frac{1 + w^2\operatorname{sech}^2\varphi_t}{w\operatorname{sech}\varphi_t}\right] d\varphi_t, \\ x_t &= \frac{s_1}{2}\left[\frac{1}{w}(\varphi_t - \varphi) + w(\tanh\varphi_t - \tanh\varphi)\right]. \end{aligned} \tag{7.5.20}$$

From (7.1.1) now consider,

$$\begin{aligned} \dot{y} &= \cos\frac{\gamma}{2}\sinh z_t = s_1 \operatorname{sech}\varphi_t \sinh\left(-s_1 s_2 \ln[w\operatorname{sech}\varphi_t]\right), \\ \dot{y} &= \frac{-s_2 \operatorname{sech}\varphi_t}{2}[e^{\ln[w\operatorname{sech}\varphi_t]} - e^{-\ln[w\operatorname{sech}\varphi_t]}], \\ dy &= \frac{-s_2 \operatorname{sech}\varphi_t}{2}\left[w\operatorname{sech}\varphi_t - [w\operatorname{sech}\varphi_t]^{-1}\right] d\varphi_t, \\ y_t &= \frac{s_2}{2}\left[\frac{1}{w}(\varphi_t - \varphi) - w(\tanh\varphi_t - \tanh\varphi)\right]. \end{aligned} \tag{7.5.21}$$

□

Integration of Horizontal Subsystem - Degenerate Cases

In the following we present the integration of horizontal subsystem in degenerate cases i.e., $\lambda \in C_4$ and $\lambda \in C_5$.

Case 4 : $\lambda \in C_4$

Theorem 7.5.4. *Integration results in case 4 are summarized as follows:*

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \operatorname{sgn}\left(\cos\frac{\gamma}{2}\right) t \\ 0 \\ 0 \end{pmatrix}. \tag{7.5.22}$$

Proof.

$$\dot{z} = \sin\frac{\gamma}{2} = \sin\left(\frac{2n\pi}{2}\right) = 0.$$

Since $z(0) = 0$,

$$z_t = 0. \tag{7.5.23}$$

Therefore,

$$\begin{aligned} \dot{x} &= \cos\frac{\gamma}{2}\cosh z = \cos\left(\frac{2n\pi}{2}\right), \\ \dot{x} &= \operatorname{sgn}\left(\cos\frac{\gamma}{2}\right), \\ x_t &= \operatorname{sgn}\left(\cos\frac{\gamma}{2}\right) t + W_x, \\ x_t &= \operatorname{sgn}\left(\cos\frac{\gamma}{2}\right) t, \end{aligned} \tag{7.5.24}$$

where $W_x = 0$ because $x(0) = 0$. Now,

$$\begin{aligned} \dot{y} &= \cos\frac{\gamma}{2}\sinh z = \cos\left(\frac{2n\pi}{2}\right)\sinh(0) = 0, \\ y_t &= W_y \\ y_t &= 0, \end{aligned} \tag{7.5.25}$$

where $W_y = 0$ because $y(0) = 0$. □

Case 5 : $\lambda \in C_5$

Theorem 7.5.5. *Integration results is case 5 are summarized as follows:*

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \operatorname{sgn}\left(\sin\frac{\gamma}{2}\right) t \end{pmatrix}. \tag{7.5.26}$$

Proof.

$$\dot{z} = \sin\frac{\gamma}{2} = \sin\left(\frac{\pi + 2n\pi}{2}\right) = \operatorname{sgn}\left(\sin\frac{\gamma}{2}\right).$$

Thus,

$$\begin{aligned} z_t &= \operatorname{sgn}\left(\sin\frac{\gamma}{2}\right) t + W_z, \\ z_t &= \operatorname{sgn}\left(\sin\frac{\gamma}{2}\right) t, \end{aligned} \tag{7.5.27}$$

where W_z=0 because $z(0) = 0$. Now,

$$\dot{x} = \cos\frac{\gamma}{2}\cosh z = \cos\left(\frac{\pi + 2n\pi}{2}\right)\cosh z = 0,$$
$$x = 0, \tag{7.5.28}$$

because $x(0) = 0$. Now,

$$\begin{aligned} \dot{y} &= \cos\frac{\gamma}{2}\sinh z = \cos\left(\frac{\pi + 2n\pi}{2}\right)\sinh z = 0, \\ y_t &= 0, \end{aligned} \tag{7.5.29}$$

because $y(0) = 0$. □

7.6 Qualitative Analysis of Projections of Extremal Trajectories on xy-Plane

The standard formula for the curvature of a plane curve $(x(t), y(t))$ is given as [74]:

$$\kappa = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}}. \tag{7.6.1}$$

Using (6.4.13),(7.6.1) curvature of projections $(x(t), y(t))$ of extremal trajectories of the Hamiltonian system (6.4.13) is given as:

$$\kappa = \frac{\sin\frac{\gamma}{2}}{\cos\frac{\gamma}{2}(\cosh 2z_t)^{\frac{3}{2}}}. \tag{7.6.2}$$

The inflection points of curves exist when $\sin\frac{\gamma}{2} = 0$ and cusps exist when $\cos\frac{\gamma}{2} = 0$ or $c = 0$. Qualitative analysis reveals that all curves $(x(t), y(t))$ have inflection points for $\lambda \in \cup_{i=1}^{3} C_i$ but only for $\lambda \in C_2$ the curves have cusps. The plots of trajectories $(x(t), y(t))$ are shown in Figures 7.3, 7.4, 7.5. In degenerate cases situation is different. The extremal trajectories q_t for $\lambda \in C_4$, are sub-Riemannian geodesics in the plane $\{z = 0\}$. The curve $(x(t), y(t))$ is a straight line on the x-axis. On the other hand, in case 5 i.e., $\lambda \in C_5$, the curve $(x(t), y(t))$ is just the initial point $(0, 0)$ for $\{x = y = 0\}$. In this case, when initial conditions $x(0) = W_x$, $y(0) = W_y$, the translations along coordinate axes are zero and the motions of pseudo Euclidean plane comprise of hyperbolic rotations only. The resulting trajectory is a quarter circle in the RS and as $t \to \infty$ the circle approaches the upper and lower arms of the hyperbola in RS of Figure 5.3.

7.7 Chapter Summary

The parametrization of extremal trajectories in the Jacobi elliptic functions is complete. We also established that there are no nontrivial abnormal extremal trajectories. This concluded

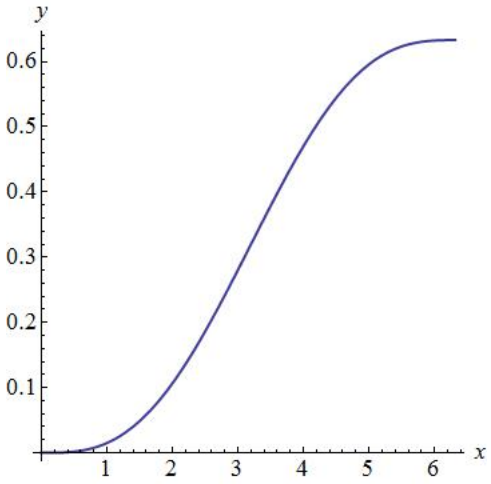

Figure 7.3: Cuspless Trajectories $\lambda \in C_1$

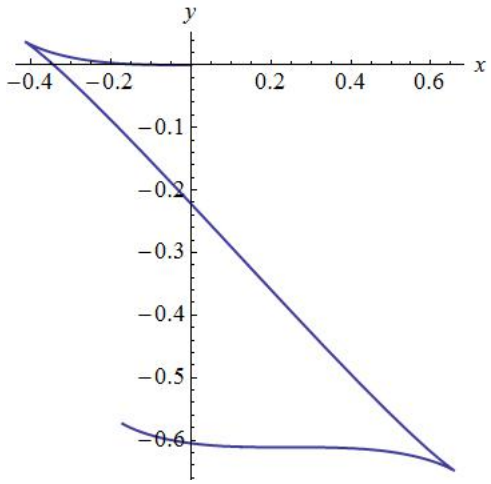

Figure 7.4: Trajectories with Cusps $\lambda \in C_2$

the research work under consideration on the Lie group SH(2). However, the solution is yet not complete without the optimality analysis. The next chapter deals with the optimality analysis based on Maxwell strata where we will use the symmetries of the vertical subsystem to find the points where extremal trajectories lose optimality.

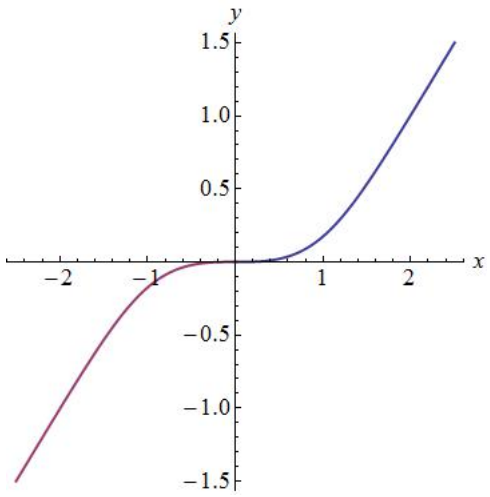

Figure 7.5: Critical Trajectories $\lambda \in C_3$

Chapter 8

Maxwell Strata

We spared great deal of effort in proving integrability of the normal Hamiltonian system and obtaining a parametrization of normal extremal trajectories in terms of convenient variables and some elementary analytical functions, yet, the rationale behind solving the nonlinear differential equations of the Hamiltonian system might not be apparent to a control engineer. Beginning our description of Maxwell strata in this chapter, relevance and significance of the vertical subsystem being a pendulum and the parametrization of extremal trajectories in terms of Jacobi elliptic functions will become clearer.

Ordinarily numerical solution of the differential equations of a dynamical system is sufficient for almost all objectives facing a control engineer viz. simulation and implementation. However, in contrast to other techniques in control theory, optimal control is concerned not only with obtaining the control function $t \mapsto u(t)$, but also with computation of trajectory-control pair $(q(.), u(.))$ [28]. In fact the original brachistochrone problem by Johann Bernoulli (that laid the foundations of calculus of variations and optimal control theory) was concerned with obtaining the curve along which time taken by a bead moving from a higher point to a lower point influenced only by its weight is minimized [28]. Pontryagin's maximum principle that forms the framework of this research does not give a unique optimal trajectory $q(t)$ joining q_0 to q_1 but a set of extremal trajectories that are candidate optimal only. Standard texts written from perspective of control engineering such as [75] do not highlight the need and procedure to obtain the unique optimal trajectory, because, for most applications it is sufficient to obtain open loop control $t \mapsto u(t)$ which guarantees that the trajectory followed by the dynamical system will be at least candidate optimal. Mathematical control theory is on the other hand more advanced and seeks a rigorous optimality analysis that is based on principle of elimination i.e., instead of computing the trajectories that are optimal, the ones that cease to be optimal for certain reason are eliminated from the set of extremal trajectories. It is known that a normal extremal trajectory ceases to be optimal for two reasons i.e., either because it meets another candidate optimal trajectory at a point called Maxwell point where both have equal sub-Riemannian distance, or because there exists a point called a conjugate point where a family of extremal trajectories has an envelope. Geometric inter-

pretation and computation of Maxwell point shall be the topic of discussion in this chapter, whereas, conjugate points will be discussed in the next chapter.

8.1 Geometric Interpretation of PMP

Undoubtedly, PMP is a triumph of great significance in Mathematical control theory, yet, it states only necessary optimality conditions. Mathematically this is equivalent to the first derivative test in elementary calculus or the Euler-Lagrange equations in calculus of variations. It is well known that a critical point or a candidate optimal point in elementary calculus is indeed optimal if it satisfies the sufficient conditions of second derivative test. The geometric interpretation of a critical point is simple in elementary calculus i.e., it is the point where the slope of the graph of the function becomes zero or the tangent line to the function becomes parallel to horizontal axis at the critical point. For optimality analysis of the geodesics obtained via PMP it suffices to remark that PMP gives the necessary optimality conditions, yet, for an inquisitive reader a geometric explanation of PMP and why does it fail to establish sufficient optimality conditions for geodesics will prove intellectually rewarding and aid further analysis.

Consider a general analytic optimal control problem on an analytic manifold M:

$$\dot{q} = f(q,u), \quad q \in M, \quad u \in U \subset \mathbb{R}^m, \tag{8.1.1}$$

$$q(0) = q_0, \quad q(t_1) = q_1, \quad t_1 \text{ is fixed}, \tag{8.1.2}$$

$$J = \int_0^{t_1} \Phi(q(t),u(t))dt \to \min, \tag{8.1.3}$$

where $f(q,u)$ is a family of vector fields, $\Phi(q,u) : M \times U \to \mathbb{R}$ is some function on M analytic in system state $q \in M$ and control parameter $u \in U$ and the definite integral J is the cost functional. The objective is to find the trajectory-control pair $(q_u(t), u(t))$ that minimizes J.

Consider the reachable (or attainable) set $\mathcal{A}_{q_0}(t)$ of the system (8.1.1)–(8.1.3) for time $t \geq 0$ from q_0 which is given as:

$$\mathcal{A}_{q_0}(t) = \{q_u(t) \quad | \quad u \in U\}.$$

Geometrically, reachable set of q_0 is the set $\mathcal{A}_{q_0}(t)$ of all those points that can be reached in time t starting from q_0. Hence any point $q(t_1)$ must lie on the boundary of the reachable set $\mathcal{A}_{q_0}(t_1)$ of point q_0 in time t_1 and therefore optimal control problems on a manifold M turn out to be the study of attainable set of some auxiliary control system on an extended state space [20], [76] $\widehat{M}$ given as:

$$\widehat{M} = \mathbb{R} \times M = \{\widehat{q} = (y,q) \mid y \in \mathbb{R}, q \in M\},$$

corresponding to an extended system:

$$\dot{\widehat{q}} = \widehat{f}(\widehat{q}, u), \qquad \widehat{q} \in \widehat{M}, \qquad u \in U,$$
$$\widehat{f}_u(\widehat{q}) = \begin{bmatrix} \Phi(q;u) \\ f_u(q) \end{bmatrix}, \qquad \widehat{q}(0) = \begin{bmatrix} y(0) \\ q(0) \end{bmatrix} = \begin{bmatrix} 0 \\ q(0) \end{bmatrix}.$$

Notice that $\Phi(q;u)$ is the integrand of the cost functional J and its initial value is zero because at time $t = 0$ there is no cost as system is at rest. Hence, trajectories of the extended system are given as:

$$\widehat{q}_u(t) = \begin{bmatrix} J_t(u) \\ q_u(t) \end{bmatrix},$$

and the corresponding attainable set of the extended system

$$\widehat{\mathcal{A}}_{(0,q_0)}(t) = \{(J_t(u), q_u(t)) \mid u \in U\}.$$

As argued earlier, the endpoint $\widehat{q}(t_1)$ must belong to the boundary of the attainable set $\widehat{\mathcal{A}}_{(0;q_0)}(t_1)$ and in order to find the optimal trajectories, the first step is to find the trajectories coming to the boundary of $\widehat{\mathcal{A}}_{(0;q_0)}(t_1)$, then the second step is to choose optimal among them [20], [76]. The first step is essentially solving the Hamiltonian system given by PMP and the second step is to eliminate the extremal trajectories that fail to be optimal at some point. Geometrically, extremal trajectories given by PMP lose optimality because the corresponding extended trajectory $\widehat{q}(t), \quad t \in [0, t_1]$ comes to the boundary of the extended attainable set $\widehat{\mathcal{A}}_{(0;q_0)}(t_1)$, but not essentially to the "lower part" of the boundary i.e., $\mathcal{A}_{q_0}(t_1) \subset M$.

As pointed out in [20], [76], the first step which is solving the Hamiltonian system is much more important than second one. Once we have a convenient parametrization and analytical solution to Hamiltonian system, the optimality analysis essentially reduces to the study of dynamics of boundary of attainable sets and that is exactly what we do in the sequel.

It is generally difficult to find the first Maxwell point. Since the vertical subsystem of the Hamiltonian system (6.5.5) is a pendulum, its phase portrait admits discrete symmetries (rotations and reflections) forming a symmetry group preserving time on the geodesics. We find the Maxwell point corresponding to these symmetries in terms of three hypersurfaces in the state space M containing all such Maxwell points. Using computer-aided symbolic computations we obtain the time of first Maxwell point corresponding to various energy levels of the pendulum. The upper bound on first Maxwell time allows us to define an upper bound on the cut time (time at which a geodesic loses optimality globally).

8.2 Discrete Symmetries of the Vertical and the Horizontal Subsystem of the Hamiltonian System

We now analyze symmetries in the vertical subsystem of the normal Hamiltonian system (6.4.13) to obtain a characterization of the Maxwell points. The analysis and organization

of this section is based on the description of the Maxwell strata given in [46], [47], [1], [77], [53], [78] with corresponding results for the problem under consideration.

Symmetries of the Vertical Subsystem

Reflection Symmetries in the Vertical Subsystem

Since the vertical subsystem of the Hamiltonian system is a mathematical pendulum (6.5.5), we exploit the reflection symmetries in the phase cylinder of the pendulum to compute the discrete symmetries of the exponential mapping. The reflection symmetries in the phase portrait of a standard pendulum are given as:

$$\begin{aligned}
&\varepsilon^1 : (\gamma, c) \to (\gamma, -c),\\
&\varepsilon^2 : (\gamma, c) \to (-\gamma, c),\\
&\varepsilon^3 : (\gamma, c) \to (-\gamma, -c),\\
&\varepsilon^4 : (\gamma, c) \to (\gamma + 2\pi, c),\\
&\varepsilon^5 : (\gamma, c) \to (\gamma + 2\pi, -c),\\
&\varepsilon^6 : (\gamma, c) \to (-\gamma + 2\pi, c),\\
&\varepsilon^7 : (\gamma, c) \to (-\gamma + 2\pi, -c).
\end{aligned} \tag{8.2.1}$$

Symmetries (8.2.1) form a symmetry group G of parallelepiped with composition as group operation and ε^i being the elements of the group. The symmetries $\varepsilon^3, \varepsilon^4, \varepsilon^7$ preserve the direction of time, however, symmetries $\varepsilon^1, \varepsilon^2, \varepsilon^5, \varepsilon^6$ reverse the direction of time [1]. As it is evident, symmetries where reflection about both axes of phase portrait occurs preserve the direction of time and others reverse the direction of time.

Reflections of Trajectories of the Pendulum

Proposition 4.1 from [1] gives the transformations that result in reflection of the phase portrait of pendulum and is reproduced here for sake of completeness.

Proposition 8.2.1. *Reflections (8.2.1) in the phase portrait of pendulum (6.5.5) are continued to the mappings ε^i that transform trajectories $\delta_s = (\gamma_s, c_s)$ of the pendulum into the trajectories $\delta_s^i = (\gamma_s^i, c_s^i)$ as follows:*

$$\varepsilon^i : \delta = \{(\gamma_s, c_s) | s \in [0, t]\} \longmapsto \delta^i = \{(\gamma_s^i, c_s^i) | s \in [0, t]\}, \quad i = 1, \ldots, 7, \tag{8.2.2}$$

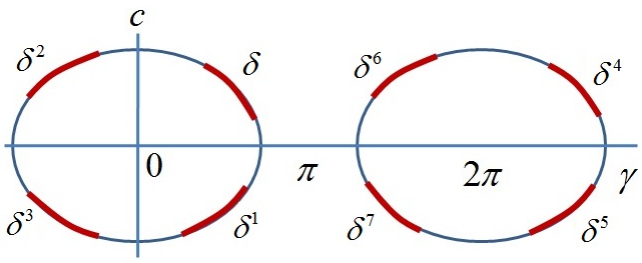

Figure 8.1: Reflections $\varepsilon^i : \delta \to \delta^i$of trajectories of a pendulum

where,

$$\begin{aligned}
\delta^1 &: (\gamma_s^1, c_s^1) = (\gamma_{t-s}, -c_{t-s}),\\
\delta^2 &: (\gamma_s^2, c_s^2) = (-\gamma_{t-s}, c_{t-s}),\\
\delta^3 &: (\gamma_s^3, c_s^3) = (-\gamma_s, -c_s),\\
\delta^4 &: (\gamma_s^4, c_s^4) = (\gamma_s + 2\pi, c_s),\\
\delta^5 &: (\gamma_s^5, c_s^5) = (\gamma_{t-s} + 2\pi, -c_{t-s}),\\
\delta^6 &: (\gamma_s^6, c_s^6) = (-\gamma_{t-s} + 2\pi, c_{t-s}),\\
\delta^7 &: (\gamma_s^7, c_s^7) = (-\gamma_s + 2\pi, -c_s).
\end{aligned} \tag{8.2.3}$$

For the instant of time $s = t/2$, the reflections of extremal trajectories $\{(\gamma_s, c_s)\} \mapsto \{(\gamma_s^i, c_s^i)\}$ given by (8.2.3) reduce to the reflections of points $\{(\gamma, c)\} \mapsto \{(\gamma^i, c^i)\}$ given by (8.2.1). In this sense we write in Proposition 8.2.1 that the reflections are continued to the mappings ε^i.

Proof. The proof of the proposition given by [1], [47] is repeated here for sake of completeness. We deal with only Case 1.

$$\dot{\gamma}_s^1 = \frac{d}{ds}\gamma_s^1 = \frac{d}{ds}\gamma_{t-s} = -\dot{\gamma}_{t-s} = -c_{t-s} = c_s^1, \tag{8.2.4}$$

$$\dot{c}_s^1 = \frac{d}{ds}c_s^1 = \frac{d}{ds}(-c_{t-s}) = \dot{c}_{t-s} = -\sin(\gamma_{t-s}). \tag{8.2.5}$$

Proof for all other cases is similar. □

Mappings (8.2.2) are shown in Figure 8.1.

Relationship between the Trajectories of the Original System and the System Transformed via Reflections

Proposition 8.2.1 bears nice intuitive interpretation. As is evident from (8.2.3) that despite action of ε^1 on the trajectories, the system equations still represent a pendulum. The phase portrait of the system obtained by reflection is same as that of pendulum, however its trajectories are different from the original system in a particular way. The term $\gamma_s^1 = \gamma_{t-s}$ implies that the integral curve of the transformed system is obtained by flipping the integral curve

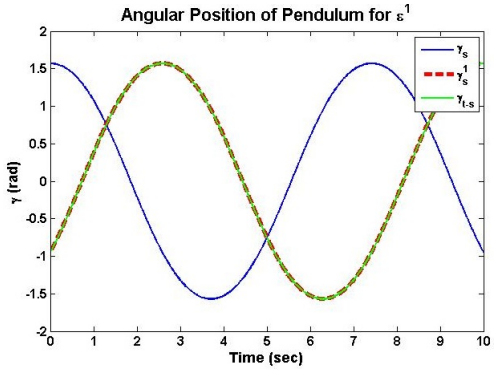

Figure 8.2: Trajectories of the System (6.5.5) under Transformation ε^1

of original system in time. The initial time $s = 0$ becomes final time $s = t$ for the reflected system and vice versa. Correspondingly the trajectory γ_s^1 is the time reversed version of γ_s. The trajectory $c_s^1 = -c_{t-s}$ is reversed in time as well as in magnitude. In Figure 8.2 the trajectories of the transformed system have been plotted by direct calculation (γ_s^1, c_s^1) as well as by flipping the original trajectories about respective axis $(\gamma_{t-s}, -c_{t-s})$. The two plots turn out to be the same.

Symmetries of the Horizontal Subsystem

Reflections of Normal Extremals

We now compute reflections of the normal extremals q_s via the exponential mapping of the vertical subsystem. The canonical projection π projects covectors from the cotangent bundle T^*M to the manifold M, i.e., $\pi : \lambda \in T^*M \mapsto q \in M$. The corresponding exponential map $\mathrm{Exp} : N \to M$ of the arc-length parametrized normal extremal trajectories for $N = C \times \mathbb{R}^+$ is given as:

$$\mathrm{Exp}(\nu) = Exp(\lambda, s) = \pi \circ e^{s\vec{H}}(\lambda) = \pi(\lambda_s) = q_s,$$

where $\nu = (\lambda, s) \in N$ and $\lambda_s = (\gamma_s, c_s, q_s)$ is the solution to the Hamiltonian system (6.4.13). We analyze the reflections of the normal extremal trajectories of the horizontal subsystem corresponding to the reflection symmetries of the vertical subsystem. Action of the group G on the normal extremals is defined as:

$$\varepsilon^i : \{\lambda_s \mid s \in [0,t]\} \mapsto \{\lambda_s^i \mid s \in [0,t]\}, \qquad i = 1, \ldots, 7. \tag{8.2.6}$$

The action ε_i of the group G on the vertical subsystem results in the reflection of trajectories of pendulum (8.2.3). The action of G on the horizontal subsystem, i.e., the trajectories q_s^i is described as follows:

Proposition 8.2.2. *The image $q_s^i = (x_s^i, y_s^i, z_s^i)$ of the normal extremal trajectory $q_s = (x_s, y_s, z_s)$, $s \in [0,t]$ under the action of reflections ε_i (8.2.6) is given as:*

(1) $$\begin{aligned} z_s^1 &= z_t - z_{t-s}, \\ x_s^1 &= \cosh z_t[x_t - x_{t-s}] - \sinh z_t[y_t - y_{t-s}], \\ y_s^1 &= \sinh z_t[x_t - x_{t-s}] - \cosh z_t[y_t - y_{t-s}]. \end{aligned}$$

(2) $$\begin{aligned} z_s^2 &= -[z_t - z_{t-s}], \\ x_s^2 &= \cosh z_t[x_t - x_{t-s}] - \sinh z_t[y_t - y_{t-s}], \\ y_s^2 &= -\sinh z_t[x_t - x_{t-s}] + \cosh z_t[y_t - y_{t-s}]. \end{aligned}$$

(3) $$\begin{aligned} z_s^3 &= -z_s, \\ x_s^3 &= x_s, \\ y_s^3 &= -y_s. \end{aligned}$$

(4) $$\begin{aligned} z_s^4 &= -z_s, \\ x_s^4 &= -x_s, \\ y_s^4 &= y_s. \end{aligned}$$

(5) $$\begin{aligned} z_s^5 &= -[z_t - z_{t-s}], \\ x_s^5 &= -\cosh z_t[x_t - x_{t-s}] + \sinh z_t[y_t - y_{t-s}], \\ y_s^5 &= \sinh z_t[x_t - x_{t-s}] - \cosh z_t[y_t - y_{t-s}]. \end{aligned}$$

(6) $$\begin{aligned} z_s^6 &= z_t - z_{t-s}, \\ x_s^6 &= -\cosh z_t[x_t - x_{t-s}] + \sinh z_t[y_t - y_{t-s}], \\ y_s^6 &= -\sinh z_t[x_t - x_{t-s}] + \cosh z_t[y_t - y_{t-s}]. \end{aligned}$$

(7) $$\begin{aligned} z_s^7 &= z_s, \\ x_s^7 &= -x_s, \\ y_s^7 &= -y_s. \end{aligned}$$

Proof. : **Case 1 - Action of** $\varepsilon^1 : (\gamma_s, c_s, q_s) \mapsto (\gamma_s^1, c_s^1, q_s^1) = (\gamma_{t-s}, -c_{t-s}, q_s^1)$

$$\begin{aligned} \dot{z}_s^1 &= \sin\frac{\gamma_s^1}{2}, \\ z_s^1 &= \int_0^s \sin\frac{\gamma_r^1}{2} dr, \\ z_s^1 &= \int_0^s \sin\frac{\gamma_{t-r}}{2} dr. \end{aligned} \tag{8.2.7}$$

Let $t - r = p \implies -dr = dp$. The limits of integral are changed as $r = 0 \implies p = t$ and $r = s \implies p = t - s$. Thus (8.2.7) becomes:

$$z_s^1 = -\left[\int_t^{t-s} \sin\frac{\gamma_p}{2} dp\right],$$
$$z_s^1 = z_t - z_{t-s}. \tag{8.2.8}$$

Similarly,

$$\begin{aligned}
\dot{x}_s^1 &= \cos\frac{\gamma_s^1}{2}\cosh z_s^1,\\
x_s^1 &= \int_0^s \cos\frac{\gamma_r^1}{2}\cosh z_r^1 dr\\
&= \int_0^s \cos\frac{\gamma_{t-r}}{2}\cosh(z_t - z_{t-r})dr\\
&= -\int_t^{t-s} \cos\frac{\gamma_p}{2}(\cosh z_t\cosh z_p - \sinh z_t \sinh z_p)dp\\
&= \cosh z_t[x_t - x_{t-s}] - \sinh z_t[y_t - y_{t-s}],
\end{aligned} \tag{8.2.9}$$

and

$$\begin{aligned}
\dot{y}_s^1 &= \cos\frac{\gamma_s^1}{2}\sinh z_s^1,\\
y_s^1 &= \int_0^s \cos\frac{\gamma_r^1}{2}\sinh z_r^1 dr\\
&= \int_0^s \cos\frac{\gamma_{t-r}}{2}\sinh(z_t - z_{t-r})dr\\
&= -\int_t^{t-s} \cos\frac{\gamma_p}{2}(\sinh z_t\cosh z_p - \cosh z_t \sinh z_p)dp\\
&= \sinh z_t[x_t - x_{t-s}] - \cosh z_t[y_t - y_{t-s}].
\end{aligned} \tag{8.2.10}$$

Proof of all other cases is similar. □

Reflections of Endpoints of Extremal Trajectories

Let us now consider the transformation of endpoints of extremal trajectories resulting from action of the reflections ε_i in the state space M:

$$\varepsilon^i : q_t \to q_t^i.$$

It can be readily seen from Proposition 8.2.2 that the point q_t^i depends only on the endpoint q_t and not on the whole trajectory $\{q_s | s \in [0,t]\}$. This is required to calculate the boundary conditions in the description of Maxwell strata corresponding to the reflection of extremal trajectories.

Proposition 8.2.3. *The action of reflections on endpoints of extremal trajectories can be defined as* $\varepsilon^i : q \mapsto q^i$, *where* $q = (x, y, z) \in M, \quad q^i = (x^i, y^i, z^i) \in M$ *and,*

$$\begin{aligned}
(x^1, y^1, z^1) &= (x\cosh z - y\sinh z,\ x\sinh z - y\cosh z,\ z),\\
(x^2, y^2, z^2) &= (x\cosh z - y\sinh z,\ -x\sinh z + y\cosh z,\ -z),\\
(x^3, y^3, z^3) &= (x,\ -y,\ -z),\\
(x^4, y^4, z^4) &= (-x,\ y,\ -z),\\
(x^5, y^5, z^5) &= (-x\cosh z + y\sinh z,\ x\sinh z - y\cosh z,\ -z),\\
(x^6, y^6, z^6) &= (-x\cosh z + y\sinh z,\ -x\sinh z + y\cosh z,\ z),\\
(x^7, y^7, z^7) &= (-x,\ -y,\ z).
\end{aligned} \tag{8.2.11}$$

Proof. : Substitute $s = t$ and $(x_0, y_0, z_0) = (0, 0, 0)$ in Proposition 8.2.2. □

Notice that Proposition 8.2.3 defines the action of reflections in the image of the exponential mapping.

Reflections as Symmetries of the Exponential Mapping

Here we calculate explicit formulas for initial values of trajectories of the pendulum corresponding to the reflections. These will be useful in characterizing the fixed points of the reflections in the preimage of the exponential map. The action of reflection in the preimage of exponential mapping is defined as:

$$\varepsilon^i : N \to N, \qquad \varepsilon^i : \nu = (\lambda, t) = (\gamma, c, t) \mapsto \nu^i = (\lambda^i, t) = (\gamma^i, c^i, t),$$

where (γ, c) are the trajectories of the pendulum with initial conditions (γ_0, c_0) and (γ^i, c^i) are the reflections of the trajectories with initial conditions (γ_0^i, c_0^i). The following proposition (a reproduction of Proposition 4.4 [1]) gives explicit formulas for (γ^i, c^i).

Proposition 8.2.4. *Let* $\nu = (\lambda, t) = (\gamma, c, t) \in N, \quad \nu^i = (\lambda^i, t) = (\gamma^i, c^i, t) \in N$. *Then,*

$$\begin{aligned}
(\gamma^1, c^1) &= (\gamma_t, -c_t),\\
(\gamma^2, c^2) &= (-\gamma_t, c_t),\\
(\gamma^3, c^3) &= (-\gamma, -c),\\
(\gamma^4, c^4) &= (\gamma + 2\pi, c),\\
(\gamma^5, c^5) &= (\gamma_t + 2\pi, -c_t),\\
(\gamma^6, c^6) &= (-\gamma_t + 2\pi, c_t),\\
(\gamma^7, c^7) &= (-\gamma + 2\pi, -c).
\end{aligned} \tag{8.2.12}$$

Proof. Substitute $s = 0$ in Proposition 8.2.1. □

Equations (8.2.11) give the explicit formulas for reflection of endpoints of the extremal trajectories in the image of exponential map, whereas, equations (8.2.12) give explicit formulas for the action of reflections ε^i on the initial points of the extremals in the preimage of the exponential mapping. The actions in M and N are both induced by reflections ε^i on extremals. Therefore it follows that the reflections ε^i for $i = 1, \ldots, 7$, are symmetries of the exponential map.

Proposition 8.2.5. *For any $\nu \in N$ and $i = 1, \ldots, 7$, we have $\varepsilon^i \circ \mathrm{Exp}(\nu) = \mathrm{Exp} \circ \varepsilon^i(\nu)$ or in other words reflection ε^i is a symmetry of the exponential mapping and the following diagram commutes:*

$$\begin{array}{ccc} N & \xrightarrow{\mathrm{Exp}} & M \\ \downarrow{\scriptstyle \epsilon^i} & & \downarrow{\scriptstyle \epsilon^i} \\ N & \xrightarrow{\mathrm{Exp}} & M \end{array} \qquad \begin{array}{ccc} \delta & \xrightarrow{\mathrm{Exp}} & q \\ \downarrow{\scriptstyle \epsilon^i} & & \downarrow{\scriptstyle \epsilon^i} \\ \delta^i & \xrightarrow{\mathrm{Exp}} & q^i \end{array}$$

8.3 Maxwell Strata Corresponding to the Reflections

Maxwell Points and Maxwell Sets

As discussed earlier an extremal trajectory can lose optimality at a Maxwell point. The term Maxwell set originates "in connection with the Maxwell rule of the van der Waals theory, according to which phase transition takes place at a value of the parameter for which two maxima of a certain smooth function are equal to each other" [79]. In optimal control theory they signify the points where the competing extremal curves with same cost functional cross each other. We now give a formal definition of Maxwell points and Maxwell strata.

Definition 8.3.1. Maxwell Point - A point q_t of a sub-Riemannian geodesic is called a Maxwell point if there exists another extremal trajectory $\tilde{q_s} \neq q_s$ such that $\tilde{q_t} = q_t$ for the instant of time $t > 0$ [1].

Definition 8.3.2. Maxwell Set/Strata - Set of all Maxwell points is called Maxwell Set and the union of all disjoint Maxwell sets is called Maxwell strata [1].

Maxwell set $\mathrm{MAX}^i, \quad i = 1, \ldots, 7$ corresponding to the reflections ε^i in the preimage of the exponential mapping $N = T^*_q M$ are defined as follows:

$$\mathrm{MAX}^i = \left\{ \nu = (\lambda, t) \in N = C \times \mathbb{R}^+ \quad | \quad \lambda \neq \lambda^i, \quad \mathrm{Exp}(\lambda, t) = \mathrm{Exp}(\lambda^i, t) \right\}. \tag{8.3.1}$$

Intuitively this definition states that Maxwell sets contain those points of M that are intersecting points of the trajectories $\mathrm{Exp}(\lambda, t)$ corresponding to covectors $\lambda \in T^*M \quad | \quad \lambda \neq$

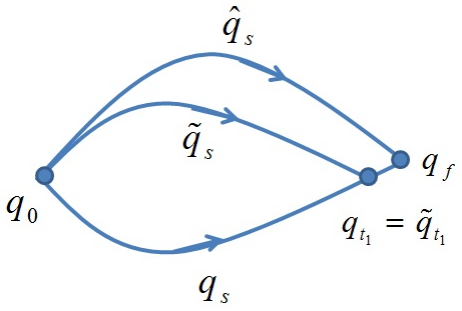

Figure 8.3: Concept of Maxwell Point - Non-optimality of the geodesic q_s after the Maxwell point q_{t_1}

$\lambda^i \quad \forall s \in [0,t]$ in T^*M i.e., the covectors are unique $\forall s \in [0,t]$. The corresponding Maxwell strata in the image of the exponential mapping are defined as:

$$\mathrm{Max}^i = \mathrm{Exp}(\mathrm{MAX}^i) \subset M. \tag{8.3.2}$$

S. Jacquet [65] proved that for an analytic problem, a trajectory cannot be optimal after a Maxwell point. We state and prove the following proposition due to S. Jacquet [65] for the sake of completeness of the exposition.

Proposition 8.3.3. *Let q_s and $\tilde{q}_s$ be two distinct geodesics: $q_s \not\equiv \tilde{q}_s : \quad s \in [0,t_1]$. If $q_{t_1} = \tilde{q}_{t_1}$, then for any $t_2 > t_1$ the geodesic $q_s, \quad s \in [0,t_2]$, is not optimal.*

In other words, for any $t_2 > t_1$ there exists a geodesic $\widehat{q}_s$, $s \in [0,t_2]$, of a smaller sub-Riemannian length than that of q_s, $s \in [0,t_2]$, connecting q_0 and q_{t_2} (see Fig 8.3). Therefore any geodesic $q_s = \mathrm{Exp}(\lambda, s)$ is non-optimal after the Maxwell time t_1, $(\lambda, t_1) \in \mathrm{MAX}$.

Proof. The proof is by contradiction. Suppose that for some$t_2 > t_1$ the geodesic $q_s, s \in [0,t_2]$, is optimal. Then the broken geodesic,

$$q'_s = \begin{cases} \tilde{q}_s, & s \in [0,t_1]\,, \\ q_s, & s \in [t_1,t_2]\,, \end{cases} \tag{8.3.3}$$

is also optimal and analytic because all the geodesics in the sub-Riemannian problem on SH(2) are analytic curves. Thus we have $q'_s \equiv q_s, \quad \forall s \in [t_1,t_2]$ and by the uniqueness theorem for analytic functions, these curves must coincide everywhere i.e., $q'_s \equiv q_s, \quad s \in [0,t_2]$. This suggests that essentially we have $\tilde{q}_s \equiv q_s, \quad s \in [0,t_1]$ and $\tilde{q}_s$ is not unique which contradicts the definition of Maxwell point q_t. □

Fixed Points of Reflections in the Image of Exponential Map

Since there are discrete symmetries of the exponential mapping, the idea is to exploit these symmetries and find the points where the trajectories arising out of symmetries meet the

normal extremal trajectory $q = (x, y, z)$. These points form the Maxwell set corresponding to the reflection symmetries. Consider the following functions:

$$R_1 = y\cosh\frac{z}{2} - x\sinh\frac{z}{2}, \qquad R_2 = x\cosh\frac{z}{2} - y\sinh\frac{z}{2}. \tag{8.3.4}$$

Consider x, y in hyperbolic coordinates:

$$x = \rho\cosh\chi, \quad y = \rho\sinh\chi.$$

Thus R_1 and R_2 read as:

$$R_1 = \rho\sinh\chi\cosh\frac{z}{2} - \rho\cosh\chi\sinh\frac{z}{2} = \rho\sinh(\chi - \frac{z}{2}),$$
$$R_2 = \rho\cosh\chi\cosh\frac{z}{2} - \rho\sinh\chi\sinh\frac{z}{2} = \rho\cosh(\chi - \frac{z}{2}).$$

Proposition 8.3.4. *Fixed points of the reflections $\varepsilon^i : q \mapsto q^i$ are given by the following conditions:*

(1) $q^1 = q \Longleftrightarrow R_1(q) = 0,$

(2) $q^2 = q \Longleftrightarrow z = 0,$

(3) $q^3 = q \Longleftrightarrow y = 0,\ z = 0,$

(4) $q^4 = q \Longleftrightarrow x = 0,\ z = 0,$

(5) $q^5 = q \Longleftrightarrow x = y = z = 0,$

(6) $q^6 = q \Longleftrightarrow R_2(q) = 0,$

(7) $q^7 = q \Longleftrightarrow x = 0,\ y = 0.$

Proof. We prove only Case (1): $q^1 = q$. The proof for all other cases is similar. From (8.2.11), $x^1 = x$ is equivalent to:

$$x\cosh z - y\sinh z = x,$$

which is equivalent to,

$$R_1\sinh\frac{z}{2} = 0. \tag{8.3.5}$$

Similarly, $y^1 = y$ is equivalent to:

$$R_1\cosh\frac{z}{2} = 0. \tag{8.3.6}$$

Eqs (8.3.5), (8.3.6) imply that $R_1 = 0$. Hence case (1) of Proposition 8.3.4 is proved. □

It can be observed readily that Max^i for $i = 3, 4, 5, 7$ form 0 or 1 dimensional manifolds contained in 2-dimensional manifolds formed by Max^i for $i = 1, 2, 6$. Thus we consider only the 2-dimensional Maxwell sets.

Fixed Points of the Reflections in the Preimage of the Exponential Map

In the previous section we considered the fixed points of reflections in M directly characterizing the Maxwell sets containing points where $q = q^i$. We now consider the fixed points in the preimage of the exponential map, i.e., the solutions to the equations $\lambda = \lambda^i$ for the proper characterization of the Maxwell points. We use the following coordinates in the phase cylinder of the pendulum for further analysis:

$$\tau = \frac{1}{2}(\varphi_t + \varphi),\ p = \frac{t}{2} \text{ when } \nu = (\lambda, t) \in N_1 \cup N_3, \tag{8.3.7}$$

$$\tau = \frac{1}{2k}(\varphi_t + \varphi),\ p = \frac{t}{2k} \text{ when } \nu = (\lambda, t) \in N_2. \tag{8.3.8}$$

Proposition 8.3.5. *Fixed points of the reflections ε^i, $i = 1, 2, 6$, in the preimage of the exponential map are given as:*

(1) $\lambda^1 = \lambda \Leftrightarrow \ \mathrm{cn}\tau = 0, \quad \lambda \in C_1,$

(2) $$\lambda^2 = \lambda \Leftrightarrow \left\{ \begin{array}{cc} \mathrm{sn}\tau = 0, & \lambda \in C_1 \cup C_2 \\ \tau = 0, & \lambda \in C_3 \end{array} \right\},$$

(3) $\lambda^6 = \lambda \Leftrightarrow \ \mathrm{cn}\tau = 0, \quad \lambda \in C_2.$

Proof. **Case 1 -** $\lambda^1 = \lambda$. It follows from Proposition 8.2.4 that if $\lambda \in C_1$, then $\lambda^i \in C_1$. Using Proposition 8.2.4,

$$\lambda^1 = \lambda \Longleftrightarrow \gamma_t = \gamma, \qquad -c_t = c. \tag{8.3.9}$$

Using elliptic coordinates (7.3.1)–(7.3.4) we have:

$$\sin\frac{\gamma}{2} = s_1 k\,\mathrm{sn}\varphi \implies \sin\frac{\gamma_t}{2} = s_1 k\,\mathrm{sn}\varphi_t \implies \sin\frac{\gamma}{2} = s_1 k\,\mathrm{sn}\varphi_t \implies \mathrm{sn}\varphi_t = \mathrm{sn}\varphi. \tag{8.3.10}$$

$$\cos\frac{\gamma}{2} = s_1 \mathrm{dn}\varphi \implies \cos\frac{\gamma_t}{2} = s_1 \mathrm{dn}\varphi_t \implies \cos\frac{\gamma}{2} = s_1 \mathrm{dn}\varphi_t \implies \mathrm{dn}\varphi_t = \mathrm{dn}\varphi. \tag{8.3.11}$$

$$\frac{c}{2} = k\,\mathrm{cn}\varphi \implies \frac{c_t}{2} = k\,\mathrm{cn}\varphi_t \implies \frac{-c}{2} = k\,\mathrm{cn}\varphi_t \implies \mathrm{cn}\varphi_t = -\mathrm{cn}\varphi. \tag{8.3.12}$$

Now from [73],

$$\mathrm{cn}\tau = \mathrm{cn}\frac{\varphi_t + \varphi}{2} = \pm\sqrt{\frac{\mathrm{cn}(\varphi_t + \varphi) + \mathrm{dn}(\varphi_t + \varphi)}{1 + \mathrm{dn}(\varphi_t + \varphi)}},$$

Consider $\mathrm{cn}(\varphi_t + \varphi) + \mathrm{dn}(\varphi_t + \varphi)$,

$$\mathrm{cn}(\varphi_t + \varphi) + \mathrm{dn}(\varphi_t + \varphi) = \frac{\mathrm{cn}\varphi_t\mathrm{cn}\varphi - \mathrm{sn}\varphi_t\mathrm{sn}\varphi\mathrm{dn}\varphi_t\mathrm{dn}\varphi}{1 - k^2\mathrm{sn}^2\varphi_t\mathrm{sn}^2\varphi} + \frac{\mathrm{dn}\varphi_t\mathrm{dn}\varphi + k^2\mathrm{sn}\varphi_t\mathrm{sn}\varphi\mathrm{cn}\varphi_t\mathrm{cn}\varphi}{1 - k^2\mathrm{sn}^2\varphi_t\mathrm{sn}^2\varphi},$$

Using (8.3.10)–(8.3.12):

$$\begin{aligned}
\mathrm{cn}(\varphi_t+\varphi)+\mathrm{dn}(\varphi_t+\varphi) &= \frac{-\mathrm{cn}^2\varphi-\mathrm{sn}^2\varphi\mathrm{dn}^2\varphi+\mathrm{dn}^2\varphi+k^2\mathrm{sn}^2\varphi\mathrm{cn}^2\varphi}{1-k^2\mathrm{sn}^2\varphi_t\mathrm{sn}^2\varphi},\\
&= \frac{-\mathrm{cn}^2\varphi+\mathrm{dn}^2\varphi\left(1-\mathrm{sn}^2\varphi\right)+\left(1-\mathrm{dn}^2\varphi\right)\mathrm{cn}^2\varphi}{1-k^2\mathrm{sn}^2\varphi_t\mathrm{sn}^2\varphi},\\
&= \frac{-\left(1-\mathrm{dn}^2\varphi\right)\mathrm{cn}^2\varphi+\left(1-\mathrm{dn}^2\varphi\right)\mathrm{cn}^2\varphi}{1-k^2\mathrm{sn}^2\varphi_t\mathrm{sn}^2\varphi},\\
\Longrightarrow\ \mathrm{cn}\tau &= 0. \qquad (8.3.13)
\end{aligned}$$

For $\lambda\in C_2^{\pm}$, we have $\lambda^1\in C_2^{\mp}$ because c inverses sign. Thus $\lambda=\lambda^1$ is impossible. Similarly if $\lambda\in C_3^{i\pm}$, we have $\lambda^1\in C_3^{i\mp}, i=0,1$ because c and γ are both inverted in sign. Hence $\lambda=\lambda^1$ is impossible. The proof for all other cases is similar. □

General Description of Maxwell Strata Generated by Reflections

Propositions 8.3.4 and 8.3.5 give the multiple points in the image and fixed points in the preimage of the exponential map respectively. We now collate the results from these propositions to give general conditions under which points $q\in M$ form part of the Maxwell sets.

Proposition 8.3.6. *For $\nu=(\lambda,t)\in\cup_{i=1}^{3}N_i$ and $q=(x,y,z)=\mathrm{Exp}\,(\nu)$,*

$$\begin{aligned}
&(1)\qquad \nu\in\mathrm{MAX}^1\Leftrightarrow\left\{\begin{array}{ll} R_1(q)=0, & \mathrm{cn}\tau\neq 0 \quad \textit{for}\,\lambda\in C_1,\\ R_1(q)=0, & \textit{for}\,\lambda\in C_2\cup C_3.\end{array}\right\}\\
&(2)\qquad \nu\in\mathrm{MAX}^2\Leftrightarrow\left\{\begin{array}{ll} z=0, & \mathrm{sn}\tau\neq 0 \quad \textit{for}\ \lambda\in C_1\cup C_2,\\ z=0, & \tau\neq 0 \quad \textit{for}\ \lambda\in C_3.\end{array}\right\}\\
&(3)\qquad \nu\in\mathrm{MAX}^3\Leftrightarrow\left\{\begin{array}{ll} R_2(q)=0, & \mathrm{cn}\tau\neq 0 \quad \textit{for}\ \lambda\in C_2,\\ R_2(q)=0, & \textit{for}\ \lambda\in C_1\cup C_3.\end{array}\right\}
\end{aligned}$$

Proof. Apply Propositions 8.3.4 and 8.3.5. □

8.4 Complete Description of the Maxwell Strata

Roots of Equations $R_i(q_t)=0$ and $z_t=0$

We now study roots of the equations $R_i(q_t)=0$ and $z_t=0$ to describe the Maxwell strata in the sub-Riemannian problem on SH(2). The idea is to obtain parametrization of the roots in terms of τ and p defined in (8.3.7)–(8.3.8). Using the addition formulas for Jacobi elliptic functions we get the following representation of the functions along extremal trajectories:

Case 1 - $\lambda \in C_1$:

$$\varphi_t = \tau + p, \quad \varphi = \tau - p, \tag{8.4.1}$$

$$\sinh z_t = s_1 \frac{2k\,\mathrm{sn}p\,\mathrm{sn}\tau}{\Delta}, \tag{8.4.2}$$

$$\sinh \frac{z_t}{2} = s_1 \frac{k\,\mathrm{sn}p\,\mathrm{sn}\tau}{\sqrt{\Delta}}, \tag{8.4.3}$$

$$\cosh \frac{z_t}{2} = \frac{1}{\sqrt{\Delta}}, \tag{8.4.4}$$

$$R_1(q_t) = \frac{2k}{1-k^2}\mathrm{cn}\tau\, f_1(p), \tag{8.4.5}$$

$$R_2(q_t) = \frac{2s_1}{1-k^2}\mathrm{dn}\tau\, f_2(p), \tag{8.4.6}$$

where $\Delta = 1 - k^2\mathrm{sn}^2 p\,\mathrm{sn}^2\tau$, $f_1(p) = \mathrm{cn}p\,\mathrm{E}(p) - \mathrm{sn}p\,\mathrm{dn}p$ and $f_2(p) = \mathrm{dn}p\,\mathrm{E}(p) - k^2\mathrm{sn}p\,\mathrm{cn}p$.

Case 2 - $\lambda \in C_2$:

$$\frac{\varphi_t}{k} = \tau + p, \; \frac{\varphi}{k} = \tau - p, \tag{8.4.7}$$

$$\sinh z_t = s_2 \frac{2k\,\mathrm{sn}p\,\mathrm{sn}\tau}{\Delta}, \tag{8.4.8}$$

$$\sinh \frac{z_t}{2} = s_2 \frac{k\,\mathrm{sn}p\,\mathrm{sn}\tau}{\sqrt{\Delta}}, \tag{8.4.9}$$

$$\cosh \frac{z_t}{2} = \frac{1}{\sqrt{\Delta}}, \tag{8.4.10}$$

$$R_1(q_t) = \frac{2s_2}{1-k^2}\mathrm{dn}\tau\, f_3(p), \tag{8.4.11}$$

$$R_2(q_t) = \frac{2k}{1-k^2}\mathrm{cn}\tau\, f_4(p), \tag{8.4.12}$$

where $f_3(p) = -\mathrm{dn}p\,\mathrm{E}(p) + p\,\mathrm{dn}p(1-k^2) + k^2\mathrm{sn}p\,\mathrm{cn}p$ and $f_4(p) = -\mathrm{cn}p\,\mathrm{E}(p) + p\,\mathrm{cn}p(1-k^2) + \mathrm{sn}p\,\mathrm{dn}p$.

Case 3 - $\lambda \in C_3$:

$$\varphi_t = \tau + p, \; \varphi = \tau - p, \tag{8.4.13}$$

$$\sinh z = 2s_1 s_2 \frac{\sinh(\tau)\sinh(p)\cosh(\tau)\cosh(p)}{\Delta}, \tag{8.4.14}$$

$$\sinh \frac{z_t}{2} = s_1 s_2 \frac{\sinh(\tau)\sinh(p)}{\sqrt{\Delta}}, \tag{8.4.15}$$

$$\cosh \frac{z_t}{2} = \frac{\cosh(\tau)\cosh(p)}{\sqrt{\Delta}}, \tag{8.4.16}$$

$$R_1(q_t) = s_2 \frac{2p - \sinh 2p}{2\sqrt{\Delta}}, \tag{8.4.17}$$

$$R_2(q_t) = s_1 \frac{2p + \sinh 2p}{2\sqrt{\Delta}}, \tag{8.4.18}$$

where $\Delta = \cosh^2\tau + \sinh^2 p$.

Proposition 8.4.1. *Let $t > 0$.*

(1) If $\lambda \in C_1$ then $z_t = 0 \iff p = 2Kn, \quad \mathrm{sn}\tau = 0.$

(2) If $\lambda \in C_2$ then $z_t = 0 \iff p = 2Kn, \quad \mathrm{sn}\tau = 0.$

(3) If $\lambda \in C_3$ then $z_t = 0 \iff p = 0, \quad \tau = 0.$

Proof. Item (1) follows from (8.4.2), item (2) from (8.4.8) and item (3) from (8.4.14). □

Proposition 8.4.2. *The function $f_1(p)$ has an infinite number of roots for any $k \in [0, 1)$ given as:*

$$p = p_1^n(k), \qquad n \in \mathbb{Z}, \tag{8.4.19}$$

$$p_1^0 = 0, \tag{8.4.20}$$

$$p_1^{-n}(k) = -p_1^n(k). \tag{8.4.21}$$

Moreover, the positive roots admit the bound:

$$p_1^n(k) \in (2nK\,,\,(2n+1)K)\,, \quad n \in \mathbb{N}, \quad k \in (0,1). \tag{8.4.22}$$

Proof. Equalities (8.4.20)–(8.4.21) follow directly from the fact that $f_1(p)$ is odd.

To prove (8.4.22) consider the function $g_1(p) = f_1(p)/\mathrm{cn}p$, which has the same roots as $f_1(p)$ and also:

$$\lim_{p\to(2n-1)K+} g_1(p) \to +\infty,$$

$$\lim_{p\to(2n+1)K-} g_1(p) \to -\infty,$$

$$g_1'(p) = -\frac{(1-k^2)\mathrm{sn}^2 p}{\mathrm{cn}^2\mathrm{p}} \leq 0.$$

Hence $g_1(p)$ is decreasing on the interval $((2n-1)K\,,\,(2n+1)K)$ approaching $\pm\infty$ on the boundaries of the interval. It follows that $g_1(p)$ and therefore $f_1(p)$ admit a unique root $p = p_1^n(k)$ in each interval $((2n-1)K\,,\,(2n+1)K)$. Since $g_1(2nK) > 0$, for $n \in \mathbb{N}$, therefore $p_1^n(k) \in (2nK, (2n+1)K)$. Plots of the functions $f_1(p)$ and $g_1(p)$ for $k = 0.9$ are given in Figure 8.4. □

Lemma 8.4.3. *The function $f_2(p)$ is positive for any $p > 0$ and $k \in (0, 1)$.*

Proof. Consider the function $g_2(p) = f_2(p)/\mathrm{dn}p$ where,

$$g_2'(p) = \frac{1-k^2}{\mathrm{dn}^2 p} > 0.$$

Since $g_2(0) = 0$ therefore $g_2(p) > 0$ and $f_2(p) > 0$ for $p > 0$. □

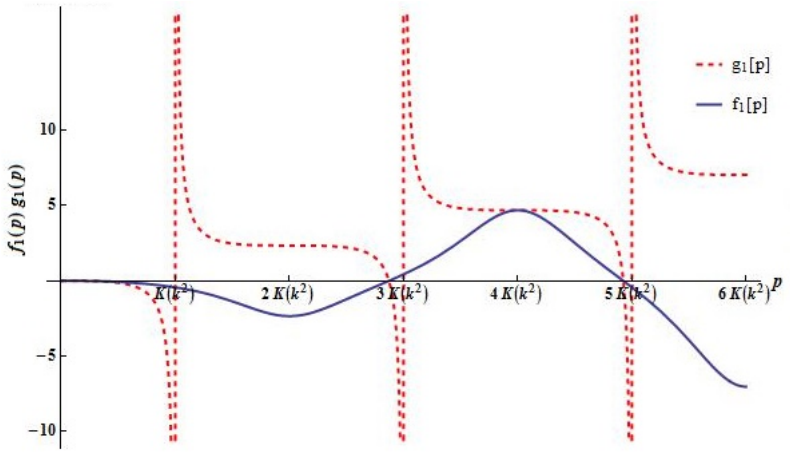

Figure 8.4: Roots of the functions $f_1(p)$ and $g_1(p)$

Lemma 8.4.4. *The function $f_3(p)$ is negative for any $p > 0$ and $k \in (0, 1)$.*

Proof. Consider the function $g_3(p) = f_3(p)/\mathrm{dn}p$ which has the same roots as $f_3(p)$ such that,

$$g_3'(p) = -\frac{(1-k^2)k^2\mathrm{sn}^2 p}{\mathrm{dn}^2 p} \leq 0.$$

Since $g_3(0) = 0$ therefore $g_3(p) < 0$ and $f_2(p) < 0$ for $p > 0$. □

Proposition 8.4.5. *The function $f_4(p)$ has an infinite number of roots for any $k \in [0, 1)$ given as:*

$$p = p_2^n(k), \qquad n \in \mathbb{Z}, \tag{8.4.23}$$

$$p_2^0 = 0, \tag{8.4.24}$$

$$p_2^{-n}(k) = -p_2^n(k). \tag{8.4.25}$$

Moreover, the positive roots admit the bound:

$$p_2^n(k) \in (2nK\,,\,(2n+1)K)\,, \quad n \in \mathbb{N}, \quad k \in (0, 1). \tag{8.4.26}$$

Proof. Equalities (8.4.24)–(8.4.25) follow directly from the fact that $f_4(p)$ is odd.

To prove (8.4.26) consider the function $g_4(p) = f_4(p)/\mathrm{cn}p$ which has the same roots as $f_4(p)$ and also:

$$\lim_{p\to(2n-1)K+} g_4(p) \to -\infty,$$

$$\lim_{p\to(2n+1)K-} g_4(p) \to +\infty,$$

$$g_4'(p) = \frac{1-k^2}{\mathrm{cn}^2 \mathrm{p}} > 0.$$

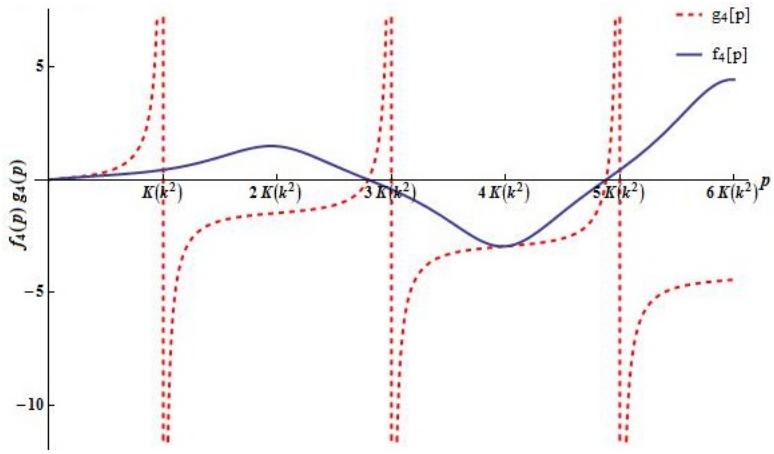

Figure 8.5: Roots of the functions $f_4(p)$ and $g_4(p)$

Hence $g_4(p)$ is increasing on the interval $((2n-1)K\,,\,(2n+1)K)$ approaching $\mp\infty$ on the boundary of the interval. It follows that $g_4(p)$ and therefore $f_4(p)$ admits a unique root $p_2^n(k)$ on each such interval. Following an argument similar to the one in Proposition 8.4.2, it follows that $p_2^n(k) \in (2nK\,,\,(2n+1)K)$. Plots of the functions $f_4(p)$ and $g_4(p)$ for $k = 0.9$ are given in Figure 8.5. □

Proposition 8.4.6. *Let $t > 0$.*

$$\text{(1)} \quad \text{If } \lambda \in C_1 \quad \text{then} \quad R_1(q_t) = 0 \quad \Longleftrightarrow \quad p = p_1^n(k) \text{ or } \mathrm{cn}\tau = 0. \tag{8.4.27}$$

$$\text{(2)} \quad \text{If } \lambda \in C_2 \quad \text{then} \quad R_1(q_t) = 0 \quad \text{is impossible.} \tag{8.4.28}$$

$$\text{(3)} \quad \text{If } \lambda \in C_3 \quad \text{then} \quad R_1(q_t) = 0 \quad \text{is impossible.} \tag{8.4.29}$$

Proof. Item (1) follows from (8.4.5) and Proposition 8.4.2. Item (2) is given from (8.4.11) and Lemma 8.4.4. Item (3) follows from (8.4.17) where $2p - \sinh 2p = 0$ for $p = 0$ and $(2p - \sinh 2p)' = 2 - 2\cosh 2p < 0$ for $p > 0$. Hence $R_1(q_t)$ does not admit any roots for $t > 0$ in this case. □

Proposition 8.4.7. *Let $t > 0$.*

$$\text{(1)} \quad \text{If } \lambda \in C_1 \quad \text{then} \quad R_2(q_t) = 0 \quad \text{is impossible.} \tag{8.4.30}$$

$$\text{(2)} \quad \text{If } \lambda \in C_2 \quad \text{then} \quad R_2(q_t) = 0, \quad \Longleftrightarrow \quad p = p_2^n(k) \text{ or } \mathrm{cn}\tau = 0. \tag{8.4.31}$$

$$\text{(3)} \quad \text{If } \lambda \in C_3 \quad \text{then} \quad R_2(q_t) = 0 \quad \text{is impossible.} \tag{8.4.32}$$

Proof. Item (1) is given from (8.4.6) and Lemma 8.4.3. Item (2) is given from (8.4.12) and Proposition 8.4.5. Item (3) follows from (8.4.18) where $2p + \sinh 2p = 0$ for $p = 0$ and $(2p + \sinh 2p)' = 2 + 2\cosh 2p > 0$ for $p \geq 0$. Hence $R_2(q_t)$ does not admit any root for $t > 0$ in this case. □

Let us now summarize the results obtained on the characterization of Maxwell strata.

Theorem 8.4.8. *The Maxwell strata* $\mathrm{MAX}^i \cap N^j$ *are given as:*

(1) $\mathrm{MAX}^1 \cap N_1 = \{\nu \in N_1 \quad | \quad p = p_1^n(k), \quad n \in \mathbb{N}, \quad \mathrm{cn}\tau \neq 0\}$,

(2) $\mathrm{MAX}^1 \cap N_2 = \mathrm{MAX}^1 \cap N_3 = \emptyset$,

(3) $\mathrm{MAX}^2 \cap N_1 = \mathrm{MAX}^2 \cap N_2 = \{\nu \in N_1 \cup N_2 \quad | \quad p = 2nK(k), \quad n \in \mathbb{N}, \quad \mathrm{sn}\tau \neq 0\}$,

(4) $\mathrm{MAX}^2 \cap N_3 = \emptyset$,

(5) $\mathrm{MAX}^6 \cap N_1 = \mathrm{MAX}^6 \cap N_3 = \emptyset$,

(6) $\mathrm{MAX}^6 \cap N_2 = \{\nu \in N_2 \quad | \quad p = p_2^n(k), \quad n \in \mathbb{N}, \quad \mathrm{cn}\tau \neq 0\}$.

Proof. This follows from the general description of the Maxwell strata and the Propositions 8.4.1, 8.4.6 and 8.4.7. □

Limit Points of the Maxwell Set

It remains to consider the points at the boundary of the Maxwell strata like the points in N_1 with $p = p_1^n(k), \quad \mathrm{cn}\tau = 0$. Since the action of reflections in the preimage of exponential map is the same for SH(2) and SE(2), it can be readily seen using Proposition 5.8 [1] that when $\nu \in N_1$, $p = p_1^1(k), \quad \mathrm{cn}\tau = 0$ and when $\nu \in N_2$, $p = p_2^1(k), \quad \mathrm{cn}\tau = 0$ then $q_t = \mathrm{Exp}(\nu)$ is a conjugate point. The same reasoning applies to the case when $\nu \in N_1, \quad \mathrm{sn}\tau = 0$ and $\nu \in N_2, \quad \mathrm{sn}\tau = 0$. Thus we get the following statement.

Proposition 8.4.9. *A point* $q_t = \mathrm{Exp}(\nu)$ *is conjugate to the initial point* q_0 *if the following conditions hold:*

(1) $\nu \in N_1, \quad p = p_1^n(k), \quad n \in \mathbb{N}, \quad \mathrm{cn}\tau = 0.$

(2) $\nu \in N_1 \cup N_2, \quad p = 2nK(k), \quad n \in \mathbb{N}, \quad \mathrm{sn}\tau = 0.$

(3) $\nu \in N_2, \quad p = p_2^n(k), \quad n \in \mathbb{N}, \quad \mathrm{cn}\tau = 0.$

Upper Bound on Cut Time

It is well known that a normal extremal trajectory cannot be optimal after the first Maxwell time. We now calculate the first Maxwell time $t_1^{\mathrm{MAX}} : C \to (0, +\infty]$.

Proposition 8.4.10. *The first Maxwell time* t_1^{MAX} *corresponding to the reflections* $\varepsilon^1, \varepsilon^2, \varepsilon^6$ *is given as:*

$$\begin{aligned} \lambda \in C_1 \quad &\Longrightarrow \quad t_1^{\mathrm{MAX}}(\lambda) = 4K(k), \\ \lambda \in C_2 \quad &\Longrightarrow \quad t_1^{\mathrm{MAX}}(\lambda) = 4kK(k), \\ \lambda \in C_3 \cup C_4 \cup C_5 \quad &\Longrightarrow \quad t_1^{\mathrm{MAX}}(\lambda) = +\infty. \end{aligned}$$

Proof. For $\lambda \in C_1, C_2, C_3$ apply Theorem 8.4.8 and Proposition 8.4.9. For $\lambda \in C_4$ and $\lambda \in C_5$, apply Theorems 7.5.4, 7.5.5 and Proposition 8.2.3. □

Using Proposition 8.4.10 we get the following global upper bound on the cut time $t_{\mathrm{cut}}(\lambda)$ for extremal trajectories.

Corollary 8.4.11. *For any $\lambda \in C$,*

$$t_{\text{cut}}(\lambda) \leq t_1^{\text{MAX}}(\lambda). \tag{8.4.33}$$

We believe that inequality (8.4.33) is in fact an equality and plan to prove this conjecture in a forthcoming work.

8.5 Symbolic Computations

It is evident that computations involved in the computations of roots of functions f_i are quite complex that are manually intractable. Hence roots of the functions f_i were calculated using Mathematica 9.0.1. The code for these computations is given in appendix B for interested reader and for sake of completeness.

8.6 Chapter Summary

Pontryagin's maximum principle gives the first order optimality conditions. The trajectories resulting from the solution of the Hamiltonian system are candidate optimal only and second order optimality conditions are checked to verify optimality. Sufficient optimality conditions are stated in terms of Maxwell points at which two trajectories with same sub-Riemannian length cross each other. In preceding paragraphs we presented complete description of Maxwell strata corresponding to the discrete symmetries of the vertical subsystem of the Hamiltonian system. We defined 2-dimensional manifolds given by functions $R_i(q)$ and z_t and proved that the these hypersurfaces contain the Maxwell points. We also stated an effective upper bound on the cut time. In next chapter we show that the cut time is indeed bounded by the first Maxwell time.

Chapter 9

Conjugate Loci

In this chapter we study local optimality of sub-Riemannian geodesics and compute the first conjugate time (i.e., the time of loss of local optimality) along extremal trajectories. Let us recall certain important facts related to conjugate points for that will also outline the scheme of further analysis. A point $q_t = \text{Exp}(\lambda, t)$ is called a conjugate point for q_0 if $\nu = (\lambda, t) = (\gamma, c, t)$ is a critical point of the exponential mapping, q_t being its critical value. In other words, this definition is given as:

$$d_\nu \text{Exp} : T_\nu N \rightarrow T_{q_t} M \text{ is degenerate,}$$

where d_νExp amounts to the Jacobian J of the exponential mapping i.e.,

$$J = \frac{\partial(x_t, y_t, z_t)}{\partial(\gamma, c, t)} = \begin{vmatrix} \frac{\partial x_t}{\partial \gamma} & \frac{\partial x_t}{\partial c} & \frac{\partial x_t}{\partial t} \\ \frac{\partial y_t}{\partial \gamma} & \frac{\partial y_t}{\partial c} & \frac{\partial y_t}{\partial t} \\ \frac{\partial z_t}{\partial \gamma} & \frac{\partial z_t}{\partial c} & \frac{\partial z_t}{\partial t} \end{vmatrix}.$$

According to the definition, roots of the equation $J = 0$ give the conjugate points and the time corresponding to these roots is called the conjugate time. Carl Gustav Jacob Jacobi (1804–1851) gave a geometric interpretation of conjugate points according to which a conjugate point q_t of a point q_0 is the point where the extremal meets the envelope of the set of extremal trajectories through q_0 [80]. This is depicted in Figure 9.1. In the local optimality analysis the first conjugate time is an important notion as this is the time at

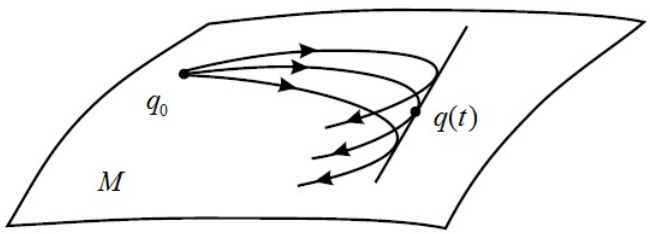

Figure 9.1: Concept of conjugate point

which an extremal trajectory loses local optimality. The first conjugate time is defined as:

$$t_1^{\text{conj}} = \inf \{t > 0 \quad | \quad t \text{ is a conjugate time along } \text{Exp}(\lambda, s), \quad s \geq 0\}.$$

Some general facts about the conjugate points are as follows [81]:

1. An instant $t > 0$ is a conjugate point if the exponential mapping for time t is degenerate i.e., the Jacobian of the exponential mapping is zero at some instant $t > 0$.

2. Morse index of the second variation of the endpoint mapping along an extremal is equal to the number of conjugate points with account of their multiplicity.

3. Morse index is equal to Maslov index of the curve in a Lagrange Grassmanian obtained by linearization of the flow of the Hamiltonian system of Pontryagin Maximum Principle.

4. Maslov index is invariant under homotopies of extremals provided that their endpoints are not conjugate.

We apply this theory for description of conjugate points in the problem (5.6.1)–(5.6.5).

9.1 Conjugate Points and Homotopy

Let us suppose $t_{\min}^{\text{conj}}$ represents the lower bound of the first conjugate time t_1^{conj}. This essentially means that there is no conjugate point $\forall t \in (0, t_{\min}^{\text{conj}})$. Hence the essential first step in estimating the bounds of first conjugate time is to prove the absence of conjugate points $\forall t \in (0, t_{\min}^{\text{conj}})$. This was achieved in Euler Elastic problem [81], SE(2) [48] and Engel group [82] via homotopy considering the fact that Maslov index (number of conjugate points along an extremal trajectory) is invariant under homotopy [83]. In order to qualify for proof of absence of conjugate points below the lower bound of first conjugate time via homotopy, the optimal control problem must satisfy a set of hypotheses **(H1)**–(**H4)** [81] outlined below.

Consider a general analytic optimal control problem on an analytic manifold M:

$$\dot{q} = f(q, u), \quad q \in M, \quad u \in U \subset \mathbb{R}^m, \tag{9.1.1}$$

$$q(0) = q_0, \quad q(t_1) = q_1, \quad t_1 \text{ is fixed}, \tag{9.1.2}$$

$$J = \int_0^{t_1} \Phi(q(t), u(t))dt \to \min, \tag{9.1.3}$$

where $f(q, u)$ is a family of vector fields and $\Phi(q, u)$ is some function on M analytic in system state $q \in M$ and control parameter $u \in U$. Note that the sub-Riemannian problem

on $M = \mathrm{SH}(2)$ (5.6.1)–(5.6.4) and (6.4.1) is of this form. Let the control dependent normal Hamiltonian of PMP for (9.1.1)-(9.1.3) be given as:

$$h_u(\lambda) = \langle \lambda, f(q,u) \rangle - \Phi(q,u). \tag{9.1.4}$$

Let a triple $(\tilde{u}(t), \lambda_t, q(t))$ represent respectively the extremal control, extremal and extremal trajectory corresponding to the normal Hamiltonian $h_u(\lambda)$. Let the following hypotheses be satisfied for (9.1.1)–(9.1.3) :

(H1) For all $\lambda \in T^*M$ and $u \in U$, the quadratic form $\frac{\partial^2 h_u}{\partial u^2}(\lambda)$ is negative definite. This is the strong Legendre condition along the extremal pair $(\tilde{u}(t), \lambda(t))$.

(H2) For any $\lambda \in T^*M$, the function $u \mapsto h_u(\lambda), \quad u \in U$, has a maximum point $\bar{u}(\lambda) \in U$:

$$h_{\bar{u}(\lambda)}(\lambda) = \max_{u \in U} h_u(\lambda), \quad \lambda \in T^*M.$$

(H3) The extremal control $\tilde{u}(.)$ is a corank one critical point of the endpoint mapping.

(H4) All trajectories of the Hamiltonian vector field $\overrightarrow{H}(\lambda), \quad \lambda \in T^*M$, are continued for $t \in [0, +\infty)$.

Under the hypotheses (**H1**)–(**H4**), the following is true for the optimal control problem of the form (9.1.1)–(9.1.3):

1. Normal extremal trajectories lose their local optimality (both strong and weak) at the first conjugate point, see [20].

2. Along each normal extremal trajectory, conjugate times are isolated one from another, see [81],[48].

We will apply the following statement for the proof of absence of conjugate points via homotopy.

Proposition 9.1.1. (Corollaries 2.2 and 2.3 [81]). *Let $(u^s(t), \lambda_t^s)$, $t \in [0, +\infty)$, $s \in [0,1]$, be continuous in parameter s family of normal extremal pairs in the optimal control problem* (9.1.1)–(9.1.3) *satisfying hypotheses* **(H1)-(H4)**. *Let $s \mapsto t_1^s$ be a continuous function, $s \in [0,1]$, $t_1^s \in (0, +\infty)$. Assume that for any $s \in [0,1]$ the instant $t = t_1^s$ is not a conjugate time along the extremal λ_t^s . If the extremal trajectory $q^0(t) = \pi(\lambda_t^0)$, $t \in (0, t_1^0]$, does not contain conjugate points, then the extremal trajectory $q^1(t) = \pi(\lambda_t^1)$, $t \in (0, t_1^1]$, also does not contain conjugate points.*

9.2 Hypothesis (H1)–(H4) for sub-Riemannian Problem on Lie Group SH(2)

Let us now check the hypothesis (**H1**)–(**H4**) for the sub-Riemannian problem on SH(2) (5.6.1)–(5.6.5).

Hypothesis (H1)–(H2)

From (6.4.9) normal Hamiltonian is given as:

$$H = h_u^{-1}(\lambda) = u_1 h_1(\lambda) + u_2 h_2(\lambda) - \frac{1}{2}\left(u_1^2 + u_2^2\right), \qquad u \in \mathbb{R}^2.$$

The Hessian of the normal Hamiltonian w.r.t to control variables is given as:

$$\frac{\partial H}{\partial u} = \begin{pmatrix} h_1 - u_1 \\ h_2 - u_2 \end{pmatrix},$$

The optimal control is given as:

$$\widetilde{u}(t) = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

For the optimal control $\widetilde{u}(t)$

$$\frac{\partial^2 H}{\partial u^2} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} < 0. \tag{9.2.1}$$

Clearly, $\frac{\partial^2 H}{\partial u^2}$ is negative definite and therefore hypothesis (**H1**) is satisfied. This in turn implies that (**H2**) also holds.

Hypothesis (H3)

Condition (**H3**) means that there exists a unique, up to a nonzero factor, extremal λ_t corresponding to the extremal control $\widetilde{u}(t)$. As outlined in [81], one easily checks that hypothesis (**H3**) is satisfied for sub-Riemannian problem on Lie group SH(2).

Hypothesis (H4)

Finally, hypothesis (**H4**) is also satisfied since the Hamiltonian vector field $\overrightarrow{H}$ is complete (its trajectories are parametrized by Jacobi's functions determined for all $t \in \mathbb{R}$).

It can be easily checked that the sub-Riemannian problem (5.6.1)–(5.6.5) satisfies the hypotheses (**H1**)–(**H4**) and therefore Proposition 9.1.1 can be used to prove bounds of the first conjugate time t_1^{conj}.

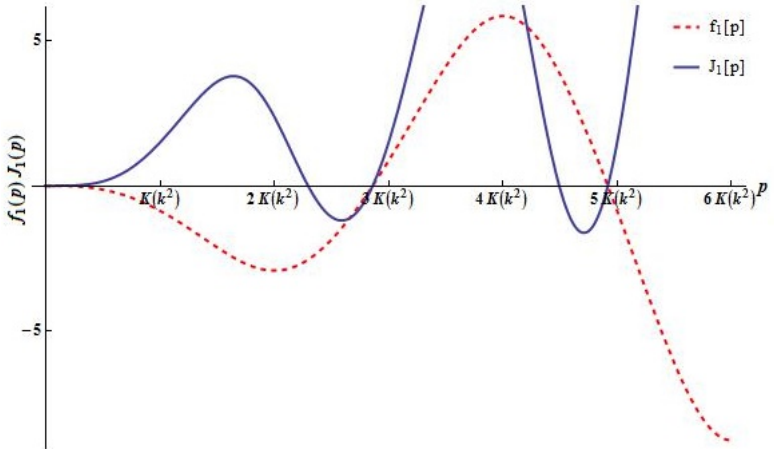

Figure 9.2: $J_1(p,\tau,k)$ and $f_1(p)$ for $k = 0.5$

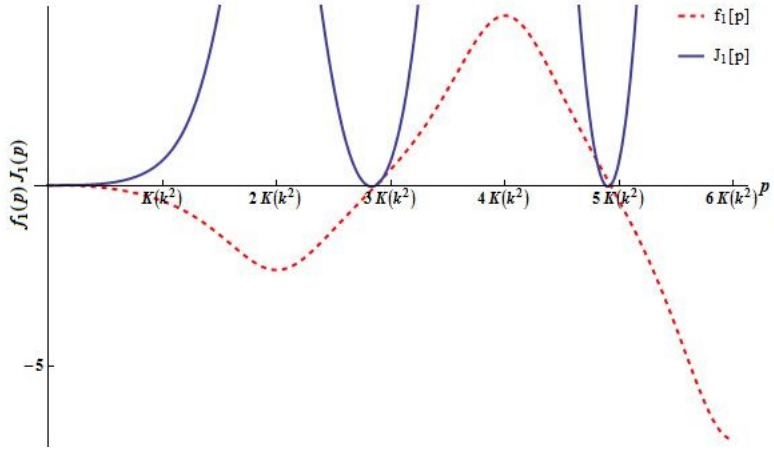

Figure 9.3: $J_1(p,\tau,k)$ and $f_1(p)$ for $k = 0.9$

9.3 Bounds of t_1^{conj} for $\lambda \in C_1$

Using the elliptic coordinates (φ, k) defined in Section 5.3.1 [3] and parametrization of extremal trajectories (7.5.1), the Jacobian of the exponential mapping is given as:

$$J = \frac{\partial(x_t, y_t, z_t)}{\partial(\varphi, k, t)} = \frac{J_1(p,\tau,k)}{(1-k^2)^2(1-k\mathrm{sn}p\,\mathrm{sn}\tau)^2}, \tag{9.3.1}$$

$$J_1(p,\tau,k) = -4k(\alpha_1+\alpha_2+\alpha_3), \tag{9.3.2}$$

$$\begin{aligned}
\alpha_1(p,\tau,k) &= \mathrm{sn}p\,\mathrm{cn}p\,\mathrm{dn}p\left(2\mathrm{E}(p) - p + k^2 p\right),\\
\alpha_2(p,\tau,k) &= -\mathrm{dn}^2 p\,\mathrm{sn}^2 p - k^2\mathrm{sn}^2 p\,\mathrm{cn}^2\tau,\\
\alpha_3(p,\tau,k) &= \mathrm{E}(p)\left(\mathrm{sn}^2 p - \mathrm{sn}^2\tau\right)\left(\mathrm{E}(p) - p + k^2 p\right),
\end{aligned}$$

where p and τ for $\lambda \in C_1$ were defined in (8.3.7). Plots of $J_1(p,\tau,k)$ are shown in Figures 9.2, 9.3.

Lemma 9.3.1. *There exists $\widehat{k} \in (0,1)$ such that for all $k \in (0,\widehat{k})$ and $p \in (0,\pi)$, the function J_1 is positive.*

Proof. The Taylor expansions of J_1 are given as:

$$J_1 = 4k \sin p(-p\cos p + \sin p), \qquad k \to 0, \tag{9.3.3}$$

$$J_1 = \frac{4}{3}kp^4 + o(k^2+p^2)^4, \qquad k^2+p^2 \to 0. \tag{9.3.4}$$

From (9.3.3) it can be readily seen that in limit passage of $k \to 0^+$, $J_1 > 0$ for $p \in (0, \pi)$. Note that $2K(0) = \pi$. Similarly from (9.3.4) it follows that $J_1 > 0$ when $k^2 + p^2 \to 0^+$. □

Lemma 9.3.2. *If $k \in (0,1)$ and $p = 2nK(k)$ for $n \in \mathbb{Z}$, then $J_1 \geq 0$.*

Proof. Direct substitution of $p = 2nK(k)$ to (9.3.2) gives:

$$J_1 = 16n^2 kE(k)\big(E(k) - (1-k^2)K(k)\big)\,\mathrm{sn}^2\tau. \tag{9.3.5}$$

Since $f(k) = E(k) - (1-k^2)K(k) > 0$ because $f(0) = 0$ and $f'(k) = kK(k) > 0$, therefore, $J_1 \geq 0$. □

Lemma 9.3.3. *The system of equations*

$$f_1(p,k) = 0, \qquad J = 0, \tag{9.3.6}$$

is incompatible for $k \in (0,1)$, $p > 0$.

Proof. We denote

$$E(u,k) = \int_0^u \sqrt{1-k^2\sin^2\varphi}\,d\varphi, \quad F(u,k) = \int_0^u \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}},$$

The system of equations (9.3.6), after the change $p = \mathrm{am}(u,k)$, turns into:

$$\begin{cases} E(u,k)\cos u = \sqrt{1-k^2\sin^2 u}\,\sin u, \\ F(u,k)\sqrt{1-k^2\sin^2 u}\,\cos u = \sin u. \end{cases} \tag{9.3.7}$$

We prove that system (9.3.7) is incompatible for $k \in (0,1)$, $u > 0$.

(1) Let $0 < u < \pi/2$. System (9.3.7) implies the equation:

$$\frac{E(u,k)}{\sqrt{1-k^2\sin^2 u}} = F(u,k)\sqrt{1-k^2\sin^2 u},$$

which is equivalent to the following equations:

$$\int_0^u \frac{\sqrt{1-k^2\sin^2\varphi}}{\sqrt{1-k^2\sin^2 u}}\,d\varphi = \int_0^u \frac{\sqrt{1-k^2\sin^2 u}}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi,$$
$$\int_0^u \left(\frac{\sqrt{1-k^2\sin^2\varphi}}{\sqrt{1-k^2\sin^2 u}} - \frac{\sqrt{1-k^2\sin^2 u}}{\sqrt{1-k^2\sin^2\varphi}}\right) d\varphi = 0,$$
$$\int_0^u \frac{1-k^2\sin^2\varphi-(1-k^2\sin^2 u)}{\sqrt{1-k^2\sin^2 u}\sqrt{1-k^2\sin^2\varphi}}\,d\varphi = 0,$$
$$\int_0^u \frac{\sin^2 u-\sin^2\varphi}{1-k^2\sin^2\varphi}\,d\varphi = 0.$$

The last equality is impossible since the function under the integral is positive for $0 < \pi < u$ (when $0 < u < \pi/2$).

(2) Equations of system (9.3.7) are violated when $\cos u = 0$ or $\sin u = 0$, i.e., at the points $u = \frac{\pi k}{2}$, $k \in \mathbb{N}$. This is checked immediately.

(3) For $\frac{\pi}{2} < u < \pi$ system (9.3.7) is incompatible since the function cos is negative, while the functions sin, E, F are positive.

(4) It remains to consider the case $u > \pi$ for $\sin u \cos u \neq 0$. In this case we multiply the equations of the system, divide the first equation by the second one, and get the following system:

$$\begin{cases} \cos^2 u\, E(u,k)F(u,k) = \sin^2 u \\ \frac{E(u,k)}{F(u,k)\sqrt{1-k^2\sin^2 u}} = \sqrt{1-k^2\sin^2 u} \end{cases} \Leftrightarrow \begin{cases} E(u,k)F(u,k) = \tan^2 u \\ E(u,k) = F(u,k)(1-k^2\sin^2 u) \end{cases}$$

The equality $1+\tan^2 u = \cos^{-2} u$ and the equation $E(u,k)F(u,k) = \tan^2 u$ imply:

$$\cos^2 u = \frac{1}{1+E(u,k)F(u,k)}.$$

Since $1-k^2\sin^2 u = 1-k^2+k^2\cos^2 u = 1-k^2+\frac{k^2}{1+E(u,k)F(u,k)}$, then the equation $E(u,k) = F(u,k)(1-k^2\sin^2 u)$ is rewritten as

$$E(u,k) = F(u,k)(1-k^2) + \frac{k^2F(u,k)}{1+E(u,k)F(u,k)}. \tag{9.3.8}$$

We have

$$E(u,k)-(1-k^2)F(u,k) = \int_0^u \left(\sqrt{1-k^2\sin^2\varphi} - \frac{1-k^2}{\sqrt{1-k^2\sin^2\varphi}}\right) d\varphi,$$
$$= \int_0^u \frac{1-k^2\sin^2\varphi-(1-k^2)}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi = \int_0^u \frac{k^2-k^2\sin^2\varphi}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi,$$
$$= k^2\int_0^u \frac{\cos^2\varphi}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi.$$

Consequently, equation (9.3.8) takes the form

$$k^2 \int_0^u \frac{\cos^2\varphi}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi = \frac{k^2F(u,k)}{1+E(u,k)F(u,k)},$$

and after dividing both sides by k^2 we get:

$$\int_0^u \frac{\cos^2\varphi}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi = \frac{F(u,k)}{1+E(u,k)F(u,k)}.$$

Since $\frac{1}{\sqrt{1-k^2\sin^2\varphi}} > 1$, $u > \pi$ then there holds the inequality

$$\frac{\cos^2\varphi}{\sqrt{1-k^2\sin^2\varphi}} > \int_0^u \cos^2\varphi\,d\varphi > \int_0^\pi \cos^2\varphi\,d\varphi = \frac{\pi}{2}$$

Consequently,

$$\frac{F(u,k)}{1+E(u,k)F(u,k)} = \int_0^u \frac{\cos^2\varphi}{\sqrt{1-k^2\sin^2\varphi}}\,d\varphi > \frac{\pi}{2}. \tag{9.3.9}$$

On the other hand, for $u \geq \pi/2$ we have $E(u,k) \geq E(k) > 1$. Consequently,

$$\frac{F(u,k)}{1+E(u,k)F(u,k)} < \frac{F(u,k)}{1+F(u,k)} < 1. \tag{9.3.10}$$

Inequalities (9.3.9) and (9.3.10) contradict one to another. This completes the proof of this lemma. □

Theorem 9.3.4. *The first conjugate time for $\lambda \in C_1$ is bounded as $4K(k) \leq t_1^{conj}(\lambda) \leq 2p_1^1(k)$. Moreover,*

$$\begin{aligned}\lim_{k\to 0^+} t_1^{conj}(\lambda) &= 2\pi,\\ \lim_{k\to 1-0} t_1^{conj}(\lambda) &= +\infty.\end{aligned}$$

Proof. We first prove the lower bound of $t_1^{\mathrm{conj}}(\lambda)$. We employ the approach adopted in the proof of Theorems 2.1, 2.2 [48] and prove that for $\lambda \in C_1$ the interval $(0, 2K(k))$ does not contain conjugate points for the extremal trajectory $q(t) = \mathrm{Exp}(\lambda, t)$.

Given any $\widehat{\lambda} \in C_1$, denote the corresponding elliptic coordinates $(\widehat{\varphi}, \widehat{k})$ and for $\widehat{t} = 4K(\widehat{k})$ denote the corresponding parameters (8.3.7) as $\widehat{p} = \widehat{t}/2$ and $\widehat{\tau} = \widehat{\varphi} + \widehat{p}$. From the discussion on conjugate points it is clear that for $p \in (0, \widehat{p})$, the extremal trajectory $\widehat{q}(t) = \mathrm{Exp}(\widehat{\lambda}, t)$ does not have conjugate points if $J_1 \neq 0$.

We choose the following family of curves in the plane (k, p) continuous in the parameter s:

$$\{(k^s, p^s) \quad | \quad s \in [0,1]\}, \qquad k^s = s\widehat{k}, \qquad p^s = 2K(k^s). \tag{9.3.11}$$

Clearly the endpoints of the curve (k^s, p^s) are $(k^0, p^0) = (0, \pi)$ and $(k^1, p^1) = (\widehat{k}, 2K(\widehat{k}))$. The corresponding family of extremal trajectories is given as:

$$q^s(t) = \mathrm{Exp}(\varphi^s, k^s, t), \quad t \in [0, t^s], \quad s \in [0, 1], \tag{9.3.12}$$

$$t^s = 2p^s, \qquad \varphi^s = \widehat{\tau} - p^s. \tag{9.3.13}$$

From Lemma 9.3.1 it is clear that for sufficiently small $s > 0$, the Jacobian $J > 0$ and hence the extremal trajectory $q^s(t)$ does not contain conjugate points for $p \in (0, 2K(k^s))$, i.e., for $t \in (0, 4K(k^s))$. Then from Proposition 9.1.1 it follows that the extremal trajectory $q^s(t)$ does not contain conjugate points for any $s \in [0, 1]$. Hence the extremal trajectory $q(t) = \mathrm{Exp}(\lambda, t), \quad \lambda \in C_1$, does not contain conjugate points in the interval $(0, 4K(k))$ and therefore $t_1^{\mathrm{conj}}(\lambda) \geq 4K(k)$.

For proof of upper bound apply Lemma 9.3.3. Hence it is proved that the first conjugate time is bounded as:

$$4K(k) \leq t_1^{\mathrm{conj}} \leq 2p_1^1(k). \tag{9.3.14}$$

From Lemma 9.3.1 and (9.3.3), the first root of J occurs at $p = \pi$ and $\lim_{k\to 0^+} 2K(k) = \pi$. Therefore,

$$\lim_{k\to 0^+} t_1^{\mathrm{conj}}(\lambda) = 4K(0) = 2\pi.$$

It can be readily seen that:

$$\lim_{k\to 1-0} t_1^{\mathrm{conj}}(\lambda) = +\infty.$$

□

Remark 9.3.5. For $\lambda \in C_1$, the instant $t = 4K(k)$ is conjugate iff $\mathrm{sn}\tau = 0$. For proof substitute $n = 1$ in (9.3.5) Lemma 9.3.2 or alternatively substitute $\mathrm{sn}\tau = 0$ in (9.3.2).

9.4 Bounds for $t_1^{\mathbf{conj}}(\lambda)$ in the Domain C_2

Using the elliptic coordinates (ψ, k) defined in Section 5.3.1 [3] and the parametrization of extremal trajectories (7.5.17), the Jacobian of the exponential mapping is given as:

$$J = \frac{\partial(x_t, y_t, z_t)}{\partial(\psi, k, t)} = \frac{-kJ_1(p, \tau, k)}{(1-k^2)^2(1-k\mathrm{sn}p\,\mathrm{sn}\tau)^2}, \tag{9.4.1}$$

where p and τ for $\lambda \in C_2$ were defined in (8.3.8) and J_1 is given by (9.3.2).

Remark 9.4.1. Notice that the Jacobian for $\lambda \in C_2$ (9.4.1) is just $-k$ times the expression of Jacobian for $\lambda \in C_1$ (9.3.1). Such a symmetry is unexpected and was not observed in equivalent problems [81],[48],[82].

Theorem 9.4.2. *The first conjugate time for $\lambda \in C_2$ is bounded as $4kK(k) \leq t_1^{conj}(\lambda) \leq 2k\, p_1^1(k)$. Moreover,*

$$\begin{aligned} \lim_{k\to 0} t_1^{conj}(\lambda) &= 2\pi k, \\ \lim_{k\to 1-0} t_1^{conj}(\lambda) &= +\infty. \end{aligned}$$

Proof. Since $J = -kJ_1$ for $\lambda \in C_2$, therefore all arguments presented in the proof of Theorem 9.3.4 apply. □

Remark 9.4.3. For $\lambda \in C_2$, the instant $t = 4k\,K(k)$ is conjugate iff $\mathrm{sn}\tau = 0$. For proof substitute $n = 1$ in (9.3.5) Lemma 9.3.2.

9.5 Conjugate Points for the Cases of Critical Energy of Pendulum

Theorem 9.5.1. (1) *If $\lambda \in C_4$, then $t_1^{\mathrm{conj}}(\lambda) = 2\pi$.*

(2) *If $\lambda \in C_3 \cup C_5$, then $t_1^{\mathrm{conj}}(\lambda) = +\infty$.*

Proof. (1) Let $\lambda \in C_4$. Take any continuous curve $\lambda^s \in C, \quad s \in [0,1]$, such that $\lambda^0 = \lambda$ and $\lambda^s \in C_1$ for $s \in (0,1]$. We have $\lim_{s\to 0+} \lambda^s = \lambda$ and $\lim_{s\to 0+} k^s = 0$, thus $\lim_{s\to 0+} t_1^{\mathrm{conj}}(\lambda^s) = 2\pi$ by Theorem 9.3.4. By continuity of the Jacobian $J(\lambda,t) = \frac{\partial q}{\partial(\lambda,t)}$, we get $J(\lambda, 2\pi) = \lim_{s\to 0+} J\left(\lambda^s, t_1^{\mathrm{conj}}(\lambda^s)\right) = 0$, thus 2π is a conjugate time along the geodesic $\mathrm{Exp}(\lambda,t)$. On the other hand, by Proposition 9.1.1, any interval $(0,\tau] \subset (0,2\pi)$ does not contain conjugate times. Consequently, $t_1^{\mathrm{conj}}(\lambda) = 2\pi$.

(2) If $\lambda \in C_3 \cup C_5$, we argue similarly. By choosing continuous curve $\lambda^s \in C, \quad s \in [0,1]$, such that $\lambda^0 = \lambda$ and $\lambda^s \in C_1$ for $s \in (0,1]$. Then $\lim_{s\to 0+} \lambda^s = \lambda$ and $\lim_{s\to 0+} k^s = 1$, thus $\lim_{s\to 0+} t_1^{\mathrm{conj}}(\lambda^s) = +\infty$ by Theorem 9.6.1. Then we get $t_1^{\mathrm{conj}}(\lambda) = +\infty$ by Proposition 9.1.1. □

Theorem 9.5.2. *The two sided bounds on $t_1^{\mathrm{conj}}(\lambda)$ for $\lambda \in C_1$ given by* Theorem 9.3.4 *are exact in the following sense:*

$$(1) \qquad \text{If } \mathrm{sn}\tau = 0 \quad \text{then} \quad t_1^{\mathrm{conj}}(\lambda) = 4K(k), \tag{9.5.1}$$

$$(2) \qquad \text{If } \mathrm{cn}\tau = 0 \quad \text{then} \quad t_1^{\mathrm{conj}} = p_1^1(k). \tag{9.5.2}$$

Proof. Substitute $\mathrm{sn}\tau = 0$ for item (1) and $\mathrm{cn}\tau = 0$ for item (2) in (9.3.2) respectively. □

Theorem 9.5.3. *The two sided bounds on $t_1^{\mathrm{conj}}(\lambda)$ for $\lambda \in C_2$ given by* Theorem 9.4.2 *are exact in the following sense:*

$$(1) \qquad \text{If } \mathrm{sn}\tau = 0 \quad \text{then} \quad t_1^{\mathrm{conj}}(\lambda) = 4kK(k), \tag{9.5.3}$$

$$(2) \qquad \text{If } \mathrm{cn}\tau = 0 \quad \text{then} \quad t_1^{\mathrm{conj}}(\lambda) = 2kp_1^1(k). \tag{9.5.4}$$

Proof. Substitute $\mathrm{sn}\tau = 0$ for item (1) and $\mathrm{cn}\tau = 0$ for item (2) in (9.3.2) respectively. □

9.6 n-th Conjugate Time

Computation of the first conjugate time is important in the study of local optimality of the extremal trajectories. It turns out that in the study of sub-Riemannian wavefront, it is essential to bound not only the first conjugate time, but all other conjugate times as well. Hence in the following, we obtain the bounds for the n-th conjugate time for $\lambda \in C_1 \cup C_2$.

Theorem 9.6.1. *The n-th conjugate time t_n^{conj} for $\lambda \in C_1$ is bounded as $4nK(k) \leq t_{2n-1}^{\mathrm{conj}} \leq 2p_1^n(k)$ and $2p_1^n(k) \leq t_{2n}^{\mathrm{conj}} \leq 4(n+1)K(k), \quad \forall n \in \mathbb{N}$.*

Proof. From lemma 9.3.2 it is readily seen that $\forall p = 2nK(k)$, the expression of the Jacobian $J_1 \geq 0$. It is trivial to see from Proposition 8.4.2 and Figure 8.4 that behavior of $f_1(p) = \mathrm{cn}p\,\mathrm{E(p)} - \mathrm{sn}p\,\mathrm{dn}p$ is uniform for all intervals $(2nK\,,\,(2n+1)K)$. Therefore at the n^{th} root $p_1^n(k)$ of the function $f_1(p)$, the Jacobian $J_1 \leq 0$. Hence the Jacobian takes values of opposite sign (or zero) at the points $p = 2nK(k)$ and $p = p_1^n(k)$. Therefore, the n^{th} conjugate time $t_n^{\mathrm{conj}}$is bounded as $4nK(k) \leq t_{2n-1}^{\mathrm{conj}} \leq 2p_1^n(k)$ and $2p_1^n(k) \leq t_{2n}^{\mathrm{conj}} \leq 4(n+1)K(k) \quad \forall n \in \mathbb{N}$. □

Corollary 9.6.2. *From* Theorem 9.4.2 *and* Theorem 9.6.1 *we see that the n-th conjugate time t_n^{conj} for $\lambda \in C_2$ is bounded as $4nkK(k) \leq t_{2n-1}^{\mathrm{conj}} \leq 2kp_1^n(k)$ and $2kp_1^n(k) \leq t_{2n}^{\mathrm{conj}} \leq 4(n+1)kK(k)$.*

Theorem 9.6.3. *The n-th conjugate times are bounded as:*

$$\begin{aligned}
\lambda \in C_1 \quad &\Longrightarrow \quad 4nK(k) \leq t_{2n-1}^{\mathrm{conj}}(\lambda) \leq 2p_1^n(k), \quad 2p_1^n(k) \leq t_{2n}^{\mathrm{conj}}(\lambda) \leq 4(n+1)K(k), \\
\lambda \in C_2 \quad &\Longrightarrow \quad 4nkK(k) \leq t_{2n-1}^{\mathrm{conj}}(\lambda) \leq 2kp_1^n(k), \quad 2kp_1^n(k) \leq t_{2n}^{\mathrm{conj}}(\lambda) \leq 4(n+1)kK(k), \\
\lambda \in C_4 \quad &\Longrightarrow \quad 2n\pi \leq t_{2n-1}^{\mathrm{conj}}(\lambda) \leq 2p_1^n(0), \quad 2p_1^n(0) \leq t_{2n}^{\mathrm{conj}}(\lambda) \leq 2(n+1)\pi.
\end{aligned}$$

Proof. The bounds follow from Theorem 9.6.1 and Corollary 9.6.2 for $\lambda \in C_1 \cup C_2$. For $\lambda \in C_4$, apply $\lim_{k\to 0+}$ to bounds for $t_n^{\mathrm{conj}}(\lambda), \quad \lambda \in C_1$. □

9.7 Numerical Calculation of Roots of Jacobian

Having found analytical results on the bound of first conjugate times for $k \in (0,1)$, we now turn towards numerical calculation of roots of Jacobian and confirmation of the bounds. For this purpose we compute the roots of Jacobian (9.3.2) for various $k \in [0.01, 0.99]$ with an

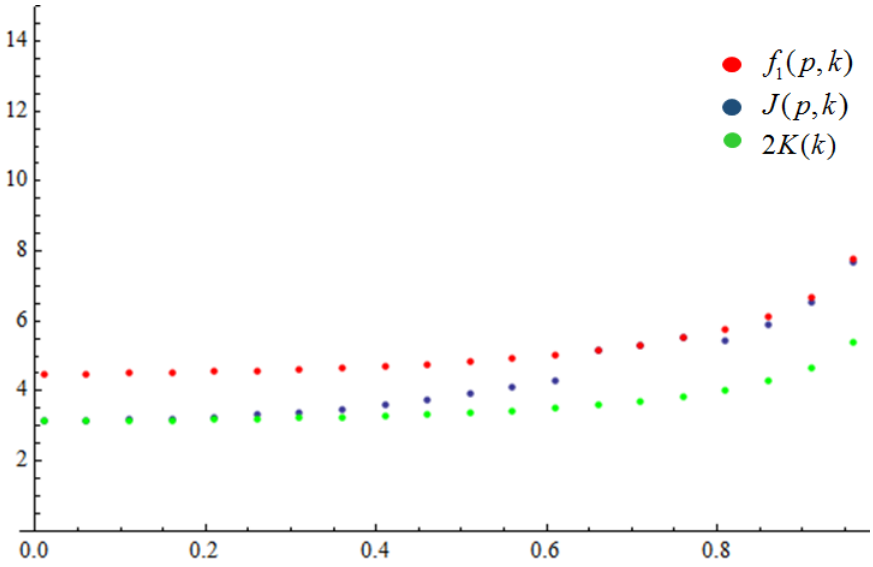

Figure 9.4: Numerical Calculation of t_1^{conj}

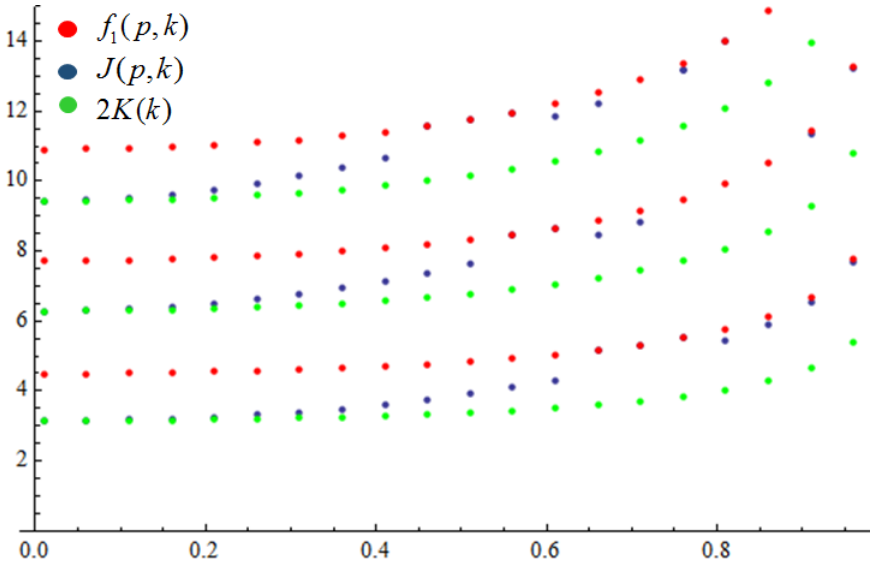

Figure 9.5: Numerical Calculation of n-th Root of Jacobian for $n = 1, 3, 5$

increment of 0.05. We also compute value of the upper and lower bound functions at these k and plot them together see Figure 9.4. Similarly, we calculate n-th root of Jacobian and plot it along with the upper and lower bound. This is presented in Figure 9.5. Mathematica code for these plots and other analytical calculations is given in appendix C.

9.8 Sub-Riemannian Sphere and Wavefront

Having explicit parametrization of the exponential mapping $\text{Exp}(\lambda, t)$, $\lambda \in C$, $t > 0$ and the global bound on the cut time, we perform a graphic study of some essential objects in the sub-Riemannian problem on SH(2) in the rectifying coordinates (R_1, R_2, z). In particular

we plot the sub-Riemannian sphere S_R and the sub-Riemannian wavefront W_R. Recall that the sub-Riemannian wavefront $W_R(q_0; R)$ at q_0 is the set of end-points of geodesics with sub-Riemannian length R starting from q_0 and the sub-Riemannian sphere $S_R(q_0; R)$ at q_0 is the set of end-points of minimizing geodesics of sub-Riemannian length R and starting from q_0:

$$\begin{aligned} W_R &= \{q = \mathrm{Exp}(\lambda, R) \in M \quad | \quad \lambda \in C\}, \\ S_R &= \{q = \mathrm{Exp}(\lambda, R) \in M \quad | \quad \lambda \in C, \quad t_{\mathrm{cut}}(\lambda) \geq R\} = \{q \in M \quad | \quad d(q_0, q) = R\}, \end{aligned}$$

where R is the radius of sub-Riemannian sphere or wavefront and $d(q_0, q_1) = \inf\{l(q(.))\}$ is the sub-Riemannian distance corresponding to sub-Riemannian length functional $l(q(.))$ (4.3.1). Note the essential difference between sub-Riemannian wavefront and sub-Riemannian sphere. The geodesics in sub-Riemannian wavefront are only locally minimizing and drawn for time greater than the cut time as well. On the contrary, the geodesics in sub-Riemannian sphere are globally minimizing and therefore drawn for time not greater than the upper bound of cut time and therefore, $S_R \subset W_R$, but $S_R \neq W_R$ for $R > 0$ and S_R is the exterior component of W_R in the following sense:

$$S_R = \partial(M \setminus W_R). \tag{9.8.1}$$

The graphic analysis based on sub-Riemannian spheres and wavefronts has tremendous importance in the solution of sub-Riemannian problems as it confirms the analytical results on Maxwell points and conjugate points. The graphic object related to the conjugate points is the sub-Riemannian caustic that is beyond the scope of this thesis. However, the plots of sub-Riemannian shpere and wavefront are based on the computation of Maxwell points and is presented below.

A plot of sub-Riemannian sphere is presented in Figure 9.6 and plots of cutout of sub-Riemannian wavefront are presented in Figures 9.7–9.8. From Figure 9.8 it is clear that the wavefront has self intersections in the planes $R_i(q_t) = 0$ and $z_t = 0$ as expected from the general and complete description of Maxwell strata. The self intersections indicate that the extremal trajectories intersect each other with same sub-Riemannian length functional. From our discussion on Maxwell points it is clear that at Maxwell points the extremal trajectories with same sub-Riemannian length intersect each other. Hence, the self intersections in sub-Riemannian wavefront signify the Maxwell points. The graphic analysis is important in the sense that it confirms our results on Maxwell points. We see that the self intersections do occur in the surfaces $R_i(q_t) = 0$ and $z_t = 0$. Figure 9.9 shows the Matryoshka of the sub-Riemannian wavefront where self intersections in wavefronts of different radii are clearly visible. The graphics again confirm the results on sub-Riemannian problems reported in literature on the nature of extremal trajectories. The extremal trajectories have infinite number of Maxwell points in the surfaces $R_i(q_t) = 0$ and $z_t = 0$. These Maxwell points

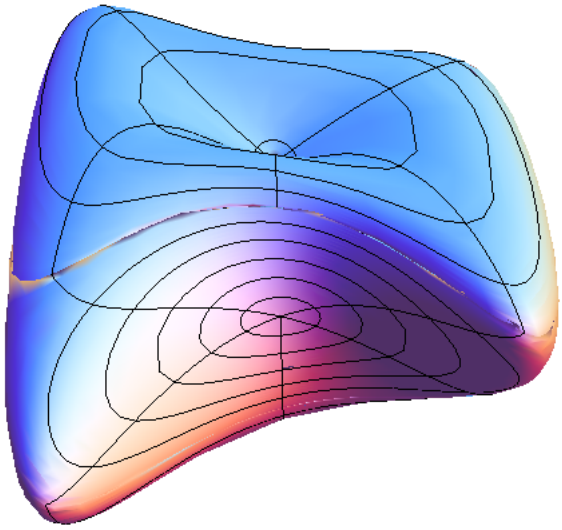

Figure 9.6: Sub-Riemannian sphere of radius 2

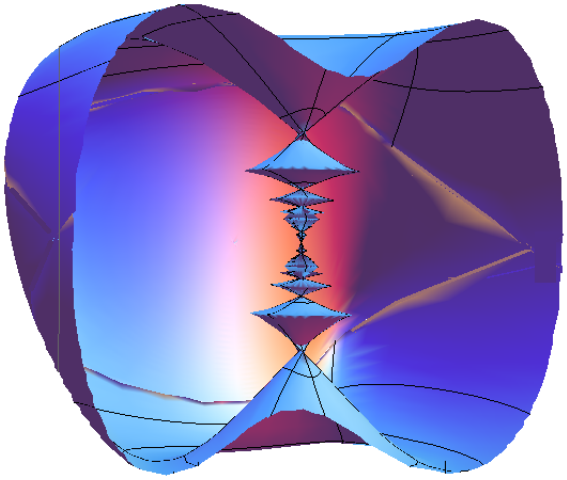

Figure 9.7: Cutout of the sub-Riemannian wavefront for $R = 2$

appear as a result of the periodic nature of the Jacobi elliptic functions that form the surfaces $R_i(q_t) = 0$ and $z_t = 0$. In Figure 9.10 we present the Matryoshka of the sub-Riemannian spheres S_R for different $R > 0$. Plots are presented from two different viewpoints for better visualization. Note that as expected, exterior view of the sub-Riemannian sphere is same as that of wavefront. This is because the sub-Riemannian sphere is the exterior component of sub-Riemannian wavefront in the sense (9.8.1). Mathematica code for these plots is given in Appendix D.

9.9 Chapter Summary

The extremal trajectories lose optimality either at Maxwell points or at the conjugate points i.e., points where the extremal trajectories have envelope. Optimality analysis presented in

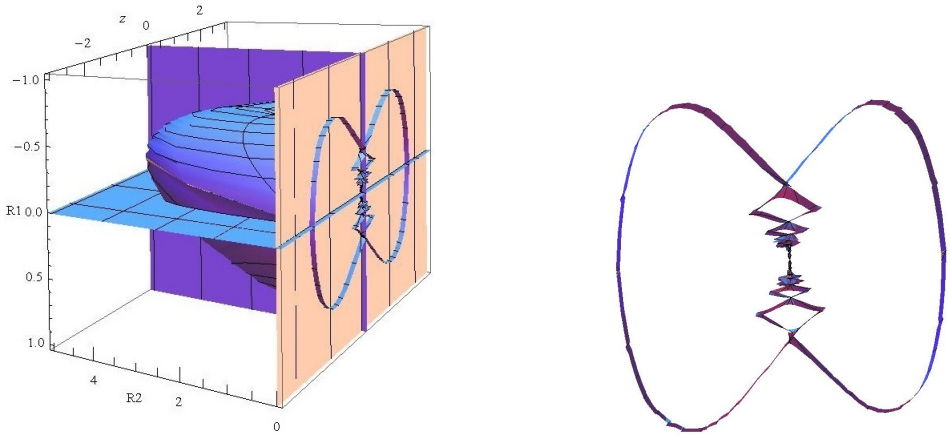

Figure 9.8: Sub-Riemannian wavefront with self intersections in the planes $R_i(q_t) = 0$ and $z_t = 0$ for $R = 2$

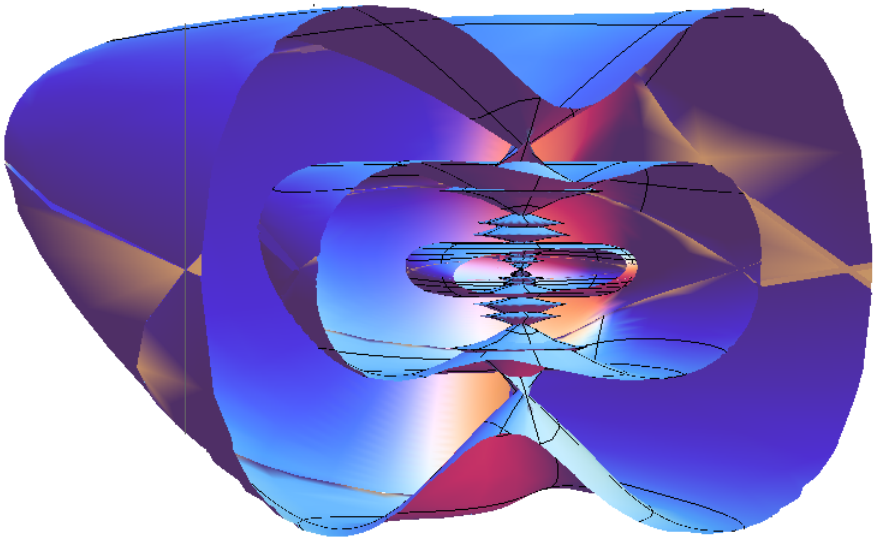

Figure 9.9: Matryoshka of sub-Riemannian wavefronts W_R for $R = 1, 2, 3$

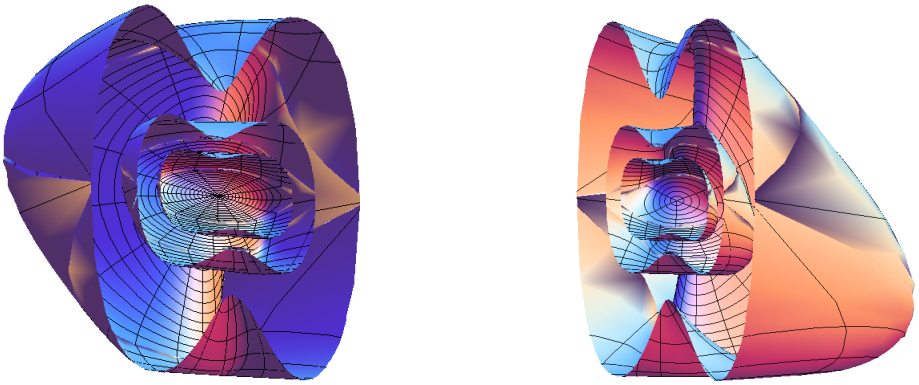

Figure 9.10: Matryoshka of sub-Riemannian spheres S_R for $R = 1, 2, 3$

this chapter has established that the time of existence of conjugate points is bounded from below by the first Maxwell time. Hence it is conjunctured that the cut time i.e., the time of loss of global optimality is the first Maxwell time. Proof of cut time equal to first Maxwell time is part of an upcoming work which is not part of this thesis.

Chapter 10

Conclusion

10.1 Main Contributions of Research Work

Pseudo Euclidean space as generalization of Euclidean space and in its own right is an important mathematical structure as well as a significant notion in relativistic mechanics. Considering a sub-Riemannian problem on pseudo Euclidean plane has important physical as well as mathematical implications as highlighted earlier. In nutshell, the main contributions of this research can be summarized as:

- ▷ Defining a sub-Riemannian problem on group SH(2).
- ▷ Proving the controllability of the control distribution.
- ▷ Considering abnormal and normal trajectories of the Hamiltonian system and proving that abnormal trajectories are not strictly abnormal.
- ▷ Proving the integrability of the dynamical system thus arguing that the ODEs are analytically solvable.
- ▷ Obtaining the transformation of sub-Riemannian problem on SH(2) to an equivalent problem on group SOLV^- thereby proving that same results can be transformed to solve equivalent problem on Lie group SOLV^-.
- ▷ Vertical subsystem depicts the dynamics of a mathematical pendulum which in turn allows integrating the system in elliptic coordinates
- ▷ Defining elliptic coordinates such that flow of the pendulum rectifies.
- ▷ Obtaining parametrization of extremal trajectories in elliptic coordinates and Jacobi elliptic functions.
- ▷ General and complete description of Maxwell strata.

- ▷ Characterization of conjugate loci.
- ▷ Global optimality analysis based on description of open dense subdomains in the preimage and the image of exponential mapping. This leads to the proof of the fact that the cut time is equal to the first Maxwell time.

As a last step the work on complete description of cut loci is in progress which is in infancy and therefore not included here. Complete characterization of cut loci will mark the completion of research on the problem under consideration.

10.2 Novelty

The work presented is novel from several perspective i.e.,

- ▷ Sub-Riemannian problem has not been considered on Lie group SH(2) prior to this work.
- ▷ In depth analysis of extremal trajectories on SH(2) was not considered earlier.
- ▷ To our knowledge, transformation between SH(2) and SOLV^- has not been obtained.
- ▷ Integrability of dynamical system is generally a difficult task and has been considered sparingly on other problems. From that viewpoint, integrability of Hamiltonian system on Lie group SH(2) is a novel contribution.
- ▷ Complete description of Maxwell strata and conjugate loci on SH(2) was never considered earlier.
- ▷ Computation of cut time for this problem is a new contribution which strengthens the techniques being developed to attack such problems.

10.3 Future Work

Important future work lays ahead from this point onwards. We have obtained such parametrization of extremal trajectories that affords good further analysis. Most natural direction of this work is to obtain the complete description of Maxwell strata thereby calculating global bound on cut time based on discrete symmetries of the vertical subsystem. To this end the methods developed in [1], [46], [47], [48] have been employed and results shall be reported in an upcoming research paper. Description of cut and conjugate loci to investigate the global and local optimality of extremal trajectories is another exciting and challenging work which is currently being targeted. The ultimate goal to be achieved in this research on Lie group

SH(2) shall be the description of global structure of exponential map and optimal synthesis. It must be emphasized though, that the future work pointed out is quite complex and challenging, suitable as the PhD thesis topic for another scholar.

10.4 Conclusion

Sub-Riemannian problems are optimal control problems with configuration manifold having a structure of a Lie group. In this work we considered the sub-Riemannian problem on Lie group SH(2). As remarked earlier, sub-Riemannian problem on SH(2) is significant both in Physics and Mathematics. Arguably, due to the natural linkage between Minkowski space-time geometry and the pseudo Euclidean plane, sub-Riemannian problem on SH(2) could potentially lead to better understanding of the relativistic mechanics. Specifically, the subgroup of hyperbolic rotations is equivalent to the Lorentz transformations.

This research led to several nontrivial achievements highlighted in the text earlier. Issues such as controllability, abnormal trajectories, integrability of Hamiltonian system and most importantly the parametrization of extremal trajectories were considered and solved. An in-depth qualitative analysis on the extremal trajectories gave further insight into the nature of sub-Riemannian geodesics. Nevertheless, important future work remains that will potentially establish the entire analysis on stronger foundations.

The applied aspect of the research is equally important. As argued, the control system physically represents a unicycle moving on a hyerbolic plane. Hence, the research results, analysis and methodology can be expected to have far reaching implications in the field of robotics where a large class of systems can be mathematically transformed into a unicycle. Since, all real world surfaces have certain curvature positive or negative, this research on motion of a unicycle on surface with constant negative curvature has obvious relevance to the motion of robot on real world surfaces. One of the important future research directions is to design a unicycle type robot that moves on a hyperbolic plane. Such application will help verify the analysis results of this research and pave the way for application of geometric control theory to real world systems.

My feelings while concluding the work on my research problem are no different than Alexander's following quote.

> Alexander wept when he heard from Anaxarchus that there was an infinite number of worlds; and his friends asked him if any accident had befallen him. He returned this answer: ***"Do you not think it a matter worthy of lamentation that when there is such a vast multitude of them, we have not yet conquered one?"***

Appendix A

Jacobi Elliptic Functions

A.1 Simply Periodic Function

Definition A.1.1. A function $f(z)$ that repeats itself after a fixed interval ω is called periodic function or simply periodic function.

$$f(z+\omega) = f(z) \tag{A.1.1}$$

where ω is called the period of the function $f(z)$. If $f(z)$ is not periodic for a submultiple of ω, then it is called fundamental period. Real world examples include current from an AC or DC generator, the motion of a pendulum or mass attached to a spring etc.

A.2 Doubly Periodic Functions

Definition A.2.1. A function $f(z)$ is called doubly periodic if there exist two constants ω_1 and $\omega_2$such that:

$$f(z+\omega_1) = f(z) = f(z+\omega_2) \tag{A.2.1}$$

for all values of complex number z [73].

The numbers $\omega_1$and ω_2 are complex and the function $f(z)$ repeats itself in two dimensions of the complex plane as against the simply periodic functions which repeat in one dimension. Jacobi in 1835 proved that the ratio of the periods $\frac{\omega_1}{\omega_2}$ cannot be real [84],[71].

A.3 Meromorphic Function

Definition A.3.1. A complex-valued function of one or more complex variables is called holomorphic function if it is complex differentiable in a neighborhood $\mathcal{U}$ of every point z in its domain D.

Definition A.3.2. A complex valued function is called meromorphic on an open set $D \subset \mathbb{C}$ that is holomorphic on all D except a set of isolated points, which are the poles for function.

Complex differentiability of holomorphic function implies that it is infinitely differentiable and equal to its own Taylor series. More often the term analytic function is synonymously used instead of holomorphic function. Analytic function is not only infinitely differentiable, it can be expressed as a convergent power series in a neighborhood $\mathcal{U}$ of every point x_0 in its domain D where x_0 can be real, complex or hypercomplex.

A.4 Elliptic Functions

History

Discovery of Elliptic functions is attributed to Norwegian mathematician Niels Henrik Abel (1802-1829) who produced mathematics of outstanding caliber in his short life. In 1825 Abel delivered a lecture at Paris explaining the double periodicity of the elliptic functions. Elliptic functions were discovered as inverse of the elliptic integrals that were encountered long before Abel. The study of elliptical integrals began in 1655 with the study of arc length of an ellipse by English mathematician John Wallis (1616-1703). In 1694 Jacob Bernoulli already familiar with elliptic integrals out of his attempt to calculate the arc length of a spiral came across another elliptic integral examining the shape of an elastic rod compressed at the ends. Such a curve is called Lemniscate figure-eight or ∞ shaped curves (comes from Latin word "lēmniscātus" meaning "decorated with ribbons", in connection with the shape a ribbon takes when tied). He showed that the arc length s of Lemniscate from 0 to x as function of a parameter t was given by the integral:

$$s = \int_0^x \frac{1}{\sqrt{1-t^4}} dt \tag{A.4.1}$$

The integral equation (A.4.1) is called Elliptic integral. The adjective elliptic is due to the fact that simplest of these integrals are encountered in measurement of the perimeter (length of arc) of an ellipse. Thus in relation to its connection with the arc length of the ellipse, these are called elliptic integrals. This can be compared to the history of development of transcendental functions which appeared as inverse functions to the solution of the integral of the kind:

$$\arcsin(x) = \int_0^x \frac{1}{\sqrt{1-t^2}} dt. \tag{A.4.2}$$

Simple explicit solutions to the integrals of algebraic expressions such as the one in equation (A.4.2) were not possible. It was established that inverse functions of the trigonometric functions that were invented from ratios of sides of right triangle were the solutions of these integrals and vice versa. The trigonometric functions were then also recognized as coordinates on the perimeter of a unit circle. These functions enlarge the class of integrals that

can be computed explicitly, albeit in the form of inverses to transcendental functions. The effort to find closed form solutions for elliptic integrals equation (A.4.1) proved futile. Yet again to address the needs of integral calculus, mathematicians ingeniously defined primitive functions for such integrals as inverse of elliptic functions (to be defined shortly) to obtain closed form solutions.

Definition A.4.1. More formally, integrals of the form $\int R(x, \sqrt{P(x)}dx$ where R is a rational function of its arguments and $P(x)$ is a polynomial of degree 3 or 4 are called elliptic integrals.

Liouville in 1833 proved that these integrals cannot be evaluated in terms of the elementary algebraic expressions or transcendental functions. Euler discovered the fundamental addition theorems for elliptic integrals extending the theory further. A. M. Legendre (1752-1833) devoted much of his career to the development of elliptic integrals. He discovered that through rational transformations elliptic integrals may be brought into three canonical types formally known as elliptic integrals of first, second and third kind given as:

$$F(\varphi, k) = \int_0^{\varphi} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}} \tag{A.4.3}$$

$$E(\varphi, k) = \int_0^{\varphi} \sqrt{1 - k^2 \sin^2 \varphi} d\varphi \tag{A.4.4}$$

$$\Pi(\varphi, n, k) = \int_0^{\varphi} \frac{d\varphi}{\left(1 + n \sin^2 \varphi\right) \sqrt{1 - k^2 \sin^2 \varphi}} \tag{A.4.5}$$

Note that these integrals contain transcendental functions of degree two and are integrable in variable φ. Once we have calculated these integrals, φ which was the original variable of interest can be obtained by inverting the functions F, E, or Π. This inverse of the elliptic integral is the elliptic function. Elliptic integrals though developed were yet to be recognized as inverse functions of some transcendental like functions and it still remained to consider them on complex domain. The predicament in defining them as such stemmed from belated recognition of existence and geometric structure of complex numbers and the fact that transcendental functions were not developed as inverse of integrals equation (A.4.2) and hence the mathematicians did not recognize their utility early on. This was overcome by Abel and Jacobi independently and almost simultaneously. Both transitioned them from real to complex domain and discovered their double periodic nature and hence defined the elliptic functions as inverse of elliptic integrals. Jacobi wondered whether there exists a meromorphic function with two fundamental real periods. He established that such a function could be constant only. However, if the condition on ratio of periods is relaxed to being complex, one would get what is called elliptic functions. The development of elliptic functions remained aloof from the applications such as pendulum which became a classic example of explaining the elliptic integrals and functions lately. Now that the elliptic functions are in correct perspective, we introduce their formal definition.

Definition A.4.2. A doubly periodic meromorphic function $f(z)$ is called elliptic function. Elliptic functions are analytic everywhere in the complex plane except at their poles.

One might wonder why such long historical background is necessary to introduce an ordinary definition. In our humble opinion, the historical perspective is equally important. Once background is established, it is easier and worthwhile to grasp the concept.
Jacobi based his theory on inverse function $\varphi = \mathrm{am}u$ instead of Legendre's integrals where:

$$u = \int_0^{\varphi} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}} \tag{A.4.6}$$

The new function am_ is called the amplitude of u for reasons that will become apparent shortly. The Jacobi elliptic functions are then single valued functions:

$$\sin\varphi = \sin \mathrm{am}u, \qquad \cos\varphi \quad = \cos \mathrm{am}u, \qquad \Delta\varphi = \Delta \mathrm{am}u = \sqrt{1 - k^2 \sin^2 \varphi} \tag{A.4.7}$$

Simpler notation for these function encountered nowadays is sn$u$, cn$u$ and dn$u$. Jacobi's elliptic functions were introduced in his famous work of 1829 titled Fundamenta nova theoriae functionum ellipticarum (New foundations of the theory of the elliptic functions). After Jacobi and Abel, Weierstrass did most important development replacing three Jacobi elliptic function by one from which all others can be obtained. This ends the interesting tale of development of elliptic functions.

Derivation

The pendulum is a classic example to introduce elliptic functions and it gives a physical explanation that why these functions are periodic in two dimensions on the complex plane. We give the derivation of the elliptic integral of first kind equation (A.4.6) and leave the derivation of other functions to the interested reader. The derivation is motivated from [84]. Consider the pendulum made of massless rod of length l and a bob of mass m making an angle θ with the vertical. The coordinate axis are pointing right and up with origin at the point where the rod is attached to the support. The equations of motion and total energy of the pendulum are given as:

$$\ddot{\theta} = -\frac{g}{l} \sin\theta \tag{A.4.8}$$

$$\mathbb{E} = \frac{1}{2} m l^2 \dot{\theta}^2 + mgl(1 - \cos\theta) \tag{A.4.9}$$

$$\dot{\theta}^2 = 2\frac{\mathbb{E}}{ml^2} - 4\frac{g}{l} \sin^2 \frac{\theta}{2} \tag{A.4.10}$$

It is well known that for small angle approximation pendulum is performing simple harmonic motion and equation (A.4.8) has solution. However, beyond small angle, the motion is not simple harmonic and the equation of simple pendulum has no longer simple solution. The total energy $\mathbb{E}$ is a constant according to law of conservation of energy. The maximum energy

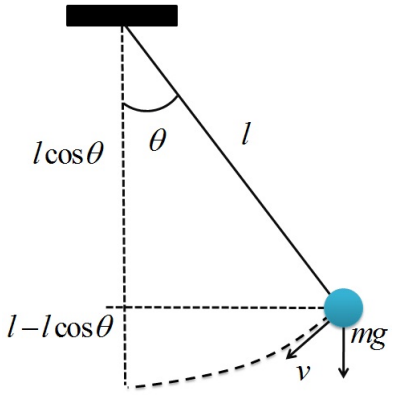

Figure A.1: Pendulum

needed to lift the pendulum from $\theta = 0$ to $\theta = \pi$ is $\mathbb{E} = 2mgl$. Hence the energy of pendulum can be given as:

$$\mathbb{E} = k^2 2mgl \tag{A.4.11}$$

for $k \geq 0$. Note that $k = 0$ implies that pendulum is at rest and $k = 1$ implies it has energy to reach the unstable equilibrium point $\theta = \pi$ though that may be reached at $t = \infty$. Consider now oscillatory motion i.e., $0 < k < 1$ with initial velocity $v = v_0$ and $\theta = 0$ at $t = 0$. The energy of the pendulum at this point is given as:

$$\mathbb{E} = \frac{1}{2}mv_0^2 \tag{A.4.12}$$

Subtracting equation (A.4.9) from equation (A.4.12) we have:

$$\begin{aligned}
v^2 &= v_0^2 - 2gl(1 - \cos\theta) \\
v^2 &= v_0^2 - 4gl\sin^2\frac{\theta}{2} \\
l^2\dot{\theta}^2 &= v_0^2 - 4gl\sin^2\frac{\theta}{2} \\
\dot{\theta}^2 &= \frac{v_0^2}{l^2} - 4\frac{g}{l}\sin^2\frac{\theta}{2} \qquad \text{(A.4.13)} \\
\dot{\theta}^2 &= \frac{4g}{l}\left(\frac{v_0^2}{4gl} - \sin^2\frac{\theta}{2}\right) \\
(d\theta)^2 &= \frac{4g(dt)^2}{l}\left(\frac{v_0^2}{4gl} - \sin^2\frac{\theta}{2}\right) \qquad \text{(A.4.14)}
\end{aligned}$$

Comparing equation (A.4.10),(A.4.11),(A.4.13) we have $k^2 = v_0^2/4gl$. Let $u = \sqrt{g/l}t$ which is the reparametrized time of motion of pendulum. Hence, $(du)^2 = \frac{g}{l}(dt)^2$ that we substitute in equation (A.4.14):

$$\left(\frac{d\theta}{du}\right)^2 = 4\left(k^2 - \sin^2\frac{\theta}{2}\right). \tag{A.4.15}$$

Let us now introduce the transformation in equation (A.4.15) to obtain the elliptic integral of first kind i.e., $k\sin\varphi = \sin\frac{\theta}{2}$.

$$\begin{aligned}
k^2\cos^2\varphi\left(\frac{d\varphi}{du}\right)^2 &= \frac{1}{4}\cos^2\frac{\theta}{2}\left(\frac{d\theta}{du}\right)^2,\\
k^2(1-\sin^2\varphi)\left(\frac{d\varphi}{du}\right)^2 &= \frac{1}{4}\cos^2\frac{\theta}{2}.4\left(k^2-\sin^2\frac{\theta}{2}\right),\\
k^2(1-\sin^2\varphi)\left(\frac{d\varphi}{du}\right)^2 &= k^2\left(1-\sin^2\varphi\right)\left(1-k^2\sin^2\varphi\right),\\
\left(\frac{d\varphi}{du}\right)^2 &= \left(1-k^2\sin^2\varphi\right),\\
u = F(\varphi,k) &= \int_0^{\varphi}\frac{d\varphi}{\left(1-k^2\sin^2\varphi\right)}. \qquad \text{(A.4.16)}
\end{aligned}$$

Note that we have obtained equation (A.4.6) which is the definition of elliptic integral of first kind. In doing so we have actually expressed reparametrized time u as a function of reparametrized pendulum position or amplitude φ. Note also that $-\pi < \theta < \pi$ is oscillatory motion, but, due to the very definition of φ its limits are $-\frac{\pi}{2} < \varphi < \frac{\pi}{2}$. The constant k is called the modulus and $k' = \sqrt{1-k^2}$ is called complementary modulus. Derivation of other elliptic functions can be seen in [84],[71]. If we introduce $\varphi = \frac{\pi}{2}$ as the limit of integration in equation (A.4.16) we get what is called the complete elliptic integral of the first kind:

$$K(k) = F(\frac{\pi}{2},k) = \int_0^{\frac{\pi}{2}}\frac{d\varphi}{\left(1-k^2\sin^2\varphi\right)}, \qquad \text{(A.4.17)}$$

$$K'(k) = K(k') = F(\frac{\pi}{2},k') = \int_0^{\frac{\pi}{2}}\frac{d\varphi}{\left(1-k'^2\sin^2\varphi\right)}. \qquad \text{(A.4.18)}$$

Note that we dropped φ as it is constant now. The integral is called complete because it corresponds to quarter period i.e. the time pendulum takes to go from $\varphi = 0$ at $x = 0$ to $\varphi = \frac{\pi}{2}$ which is the swing of pendulum for $-\frac{\pi}{2} < \varphi < \frac{\pi}{2}$ is the quarter of the total time period.

A.5 Double Periodicity

- ▷ Elliptic functions are doubly periodic. Specifically, $\mathrm{sn}(z+4K,k) = \mathrm{sn}(z,k)$ and $\mathrm{sn}(z+2iK';k) = \mathrm{sn}(z;k)$
- ▷ The imaginary period corresponds to the fact that if we replace gravity g by $-g$ we essentially have a pendulum moving upside down and imaginary period is the period of this pendulum.

Whittaker [71] proved an important theorem for dynamical system that is subjected to constraints. It states that if the constraints on the dynamical system are independent of the time of motion and also independent of forces and they are dependent only upon the position of the particles, then for such a system the integrals of the equations of motion are still real if time t and initial conditions $q_{10}, \ldots q_{n0}$ are replaced respectively by $\mathrm{i}t$ and $\mathrm{i}q_1, \ldots, \mathrm{i}q_n$. The resulting equations represent the motion which the same system would have if, with the same initial conditions, it were acted on by the same forces reversed in direction.

A.6 Identities

The identities used in this work have been stated where used.

Appendix B

Mathematica Code for Roots of Function $R_1(q_t) = 0$ for $\lambda \in C_1$

As is evident, mathematical computations in this work are complex and intractable manually. Thanks to symbolic computation software like Mathematica, it was possible to carry out our analysis and obtain analytical results [85], [86]. It however turns out that in order to obtain desired results, software has to be pushed beyond its documented capabilities. Another important step in this process is the manual analysis at each step and introduce manual simplifications wherever possible. Patience, keen observation, trial and error are essential ingredients for getting desired output for problems of this scale as the process may span over weeks to months depending upon expertise in mathematics and Mathematica. Consequently, the code is suitably commented for someone new to symbolic computations in Mathematica. Mathematica code that was written to find roots of $R_1(q_t) = 0$ for $\lambda \in C_1$ is given in the succeeding paragraphs. The code was written and tested in Mathematica 9.0.1 on Windows 8.1.

As already stated, purpose of this code is to find the roots of the function $R_1(q_t) = y_t \cosh z_t/2 - x_t \sinh z_t/2$ given in (8.3.4). Note that the extremal trajectories $q_t = (x_t, y_t, z_t)$ (7.5.1) and therefore original function $R_1(q_t)$ is parametrized in terms of φ and k (7.3.1)–(7.3.4). On the contrary we need to find the roots in terms of time which will give the first Maxwell time t_1^{MAX}. Recall that φ is the reparametrized time of motion and $\varphi_t = \varphi + t$ and therefore it is possible to reparametrize the extremal trajectories in terms of time. This reparametrization was done in (8.3.7) by virtue of which the extremal trajectories and function $R_1(q_t)$ are parametrized in terms of τ and $p = t/2$. After the reparametrization, it is necessary to simplify the resulting expressions and find the roots of $R_1(p, \tau) = 0$ in terms of p.

The code is divided into five sections. In Section B.1 we give the setup code including the definition of addition formulas for elliptic functions. We also give formulas for the functions (x_t, y_t) and $R_1(q_t) = y_t \cosh z_t/2 - x_t \sinh z_t/2$. Note that reparametrization of $\cosh \frac{z_t}{2}$ and

$\sinh \frac{z_t}{2}$ was obtained in (8.4.3)–(8.4.4) which is used in defining the function $R_1(q_t)$. In section 2 we reduce the function $R_1(q_t)$ to an equivalent but simpler form i.e., $R_1(q_t) = \frac{2k}{1-k^2}\text{cn}\tau\, f_1(p)$ where $f_1(p)$ is given in (8.4.5). In section 3 we validate the simplification process by subtracting the original expression from the simplified one. In section 4 we plot the simplified function, obtain its roots numerically to confirm its bounds and also analyze the function $g_1(p,k) = \frac{f_1(p)}{\text{cn}p}$. In section 5 we append elliptic functions interconversion rules list which is used in simplification of our results. The pseduo code for this corresponding Mathematica code is given as:

1. Initialize notebook directory
2. Define the addition / subtraction formulas for Jacobi elliptic functions
3. Define transformation rules / identities for Jacobi elliptic functions
4. Define parametric equations of extremal trajectories and the function $R_1(q)$
5. Simplify the function using builtin Mathematica functions and transformation rules
6. Obtain the roots of functions $R_1(q)$ from simplified expression
7. Plot the original and simplified function for comparison

B.1 Setup

```
SetDirectory[NotebookDirectory[]];
color := {RGBColor[1, 0, 0], RGBColor[0, 0, 1], RGBColor[0, 1, 0], RGBColor[1, 1, 0],
RGBColor[0, 1, 1]}; (*Set PlotStyle and Parametric PlotStyle to color*)
SetOptions[Plot, PlotStyle -> color];
SetOptions[ParametricPlot, PlotStyle -> color];
(*Elliptic Functions*)
amm[phi_, k_] := JacobiAmplitude[phi, k^2]
EE[phi_, k_] := EllipticE[amm[phi, k], k^2]
K[k_] := EllipticK[k^2]
EEE[k_] := EllipticE[k^2]
sn[p_, k_] := JacobiSN[p, k^2]
cn[p_, k_] := JacobiCN[p, k^2]
dn[p_, k_] := JacobiDN[p, k^2]
(*Addition formulas*) (*We first unprotect a function to modify its definition. In the end \
we again protect*)
Unprotect[JacobiDN]
(*dn(p+tau)*)
```

```
JacobiDN[p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] - k^2*JacobiSN[p, k^2]*JacobiCN[p,
k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*dn(-p+tau)*)
JacobiDN[-p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] + k^2*JacobiSN[p,
k^2]*JacobiCN[p, k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[t
k^2]^2)
Protect[JacobiDN]

Unprotect[EllipticE]
(*E(p+tau)*)
EllipticE[JacobiAmplitude[p + tau, k^2], k^2] := EllipticE[JacobiAmplitude[p, k^2], k^2] +
EllipticE[JacobiAmplitude[tau, k^2], k^2] - k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[p
+ tau, k^2]
(*E(-p+tau)*)
EllipticE[JacobiAmplitude[-p + tau, k^2], k^2] := -EllipticE[JacobiAmplitude[p, k^2], k^2]
+ EllipticE[JacobiAmplitude[tau, k^2], k^2] + k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[-
p + tau, k^2]
Protect[EllipticE]

Unprotect[JacobiSN]
(*sn(p+tau)*)
JacobiSN[p + tau, k^2] := (JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] + Ja-
cobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*sn(-p+tau)*)
JacobiSN[-p + tau, k^2] := (-JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] +
JacobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
Protect[JacobiSN]

Unprotect[JacobiCN]
(*cn(p+tau)*)
JacobiCN[p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] - JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*cn(-p+tau)*)
JacobiCN[-p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] + JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
Protect[JacobiCN]
```

```
<< "rules.txt"        (* Rules for expansion of trigonometric identities. This file is given at the end of this code *)
delta[p_, t_, k_] := 1 - k^2*sn[t, k]^2*sn[p, k]^2;
phit[p_, t_] := t + p;
phi0[p_, t_] := t - p;
w[p_, t_, k_] := 1/{dn[phi0[p, t], k] - k*cn[phi0[p, t], k]}; k0 = 0.1;
x11[p_, t_, k_] := w[p, t, k] + 1/{{1 - k^2}*w[p, t, k]};
x12[p_, t_, k_] := k*{-w[p, t, k] + 1/{{1 - k^2}*w[p, t, k]}};
y11[p_, t_, k_] := w[p, t, k] - 1/{{1 - k^2}*w[p, t, k]};
y12[p_, t_, k_] := k*{w[p, t, k] + 1/{{1 - k^2}*w[p, t, k]}};
xya[p_, t_, k_] := EE[phit[p, t], k] - EE[phi0[p, t], k];
xyb[p_, t_, k_] := sn[phit[p, t], k] - sn[phi0[p, t], k]; (* x=s1/2*{x_{11}*xya+x_{12}*xyb}
y=1/2*{y_{11}*xya-y_{12}*xyb} *)
x[p_, t_, k_] := 1/2*{x11[p, t, k]*xya[p, t, k] + x12[p, t, k]*xyb[p, t, k]};
y[p_, t_, k_] := 1/2*{y11[p, t, k]*xya[p, t, k] - y12[p, t, k]*xyb[p, t, k]};
```

B.2 Derivation of the Factorization $R_1(q_t) = \frac{2k}{1-k^2}\mathrm{cn}\tau\, f_1(p)$

```
R1 = y[p, tau, k] - x[p, tau, k] k sn[p, k] sn[tau, k];        (*This expression results from substituting values of cosh z_t/2 and sinh z_t/2 in R_1(q_t)*)
% /. tocsdErule1;        (*Apply elliptic functions rules defined in rules.txt on previous expression.*)
Together[%];
Numerator[%] ;
x1 = Denominator[%%];        (* Separate denominator and call it x1 *)
FullSimplify[%%];        (* Simplify the numerator using FullSimplify *)
Expand[%] /. ctosrules1 /. dtosrules1;        (* Expand the simplified expression and replace the cn and dn functions by sn functions for further simplification *)
FullSimplify[%];
(*This is the output of previous FullSimplify. We note a common factor and remove it for simplification.
{{{{-2 k (-1 + k^2 s^2 ts^2) (s ((1 - k^2 s^2) tc td + c d k (-1 + ts^2)) + E1 (-c d tc td + k (-1 + s^2) (-1 + ts^2)))}}}}*)
Factor[%/(2 k (1 - k^2 s^2 ts^2))];
Expand[%] /. dtosrules1 /. stocrules1;
FullSimplify[%]
{{{{tc (-(-1 + k^2) s td - c d (k s tc + E1 td) + c^2 k (E1 tc + k s td))}}}}
Factor[% /tc];
Expand[%] /. dtosrules1 /. ctosrules1;
```

```
FullSimplify[%];
(*This is the final simplified expression.
{{{{-E1 (k (-1 + s^2) tc + c d td) + s (td - k (c d tc + k s^2 td))}}}}*)
(k c tc - d td) (c E1 - s d) - %;       (*Manual simplification results in a further simpler form
which is first part of this expression. We subtract previous expression from it to verify.*)
Expand[%] /. dtosrules1 /. ctosrules1;
FullSimplify[%];
(*Final result is 0 which means our manual simplification was correct.
{{{{0}}}}*)
```

B.3 Validation

```
(* verification of the factorization R1f = 2 k/(1-k^2) tc (c E1 - s d) *)
R1 = y[p, tau, k] - x[p, tau, k] k sn[p, k] sn[tau, k];
% /. tocsdErule1;
x2 = 2 k/(1 - k^2) tc (c E1 - s d);
Together[%% - %];       (*Subtract original and simplified expression and simplify. Result
should be 0*)
Expand[%] /. ctosrules1 /. dtosrules1;
FullSimplify[%]
{{{{0}}}}
```

B.4 Plot and Numerical Calculation of Roots

```
(* Proof of the identity g1'(p) = -((1-k^2)sn^2 p)/cn^2 p *)
(c E1 - s d) /. fromcsdErule1;
c /. fromcsdErule1;
D[%%/%, p] %^2 // FullSimplify;       (*Take derivative, then consider numerator only*)
% /. tocsdErule1 /. ctosrules /. dtosrules;
FullSimplify[%] ;
%% + (1 - k^2) s^2 // Expand;       (*Now validate whether difference of simplified expres-
sion our derivative value we calculated is 0*)
% /. dtosrules /. ctosrules;
FullSimplify[%];

(*Numerical calculation of root of simplified expression of R1*)
pR1[k_] := p /. FindRoot[ cn[p, k] EE[p, k] - sn[p, k] dn[p, k] == 0, {p, K[k], 3 K[k]}]

(*Plot the function f1(p) and g1(p)*)
<< PlotLegends`       (*Include PlotLegends package*)
```

```
k0 := 0.9;
tic = {{{K[k0], K[k]}, {2 K[k0], 2 K[k]}, {3 K[k0], 3 K[k]}, {4 K[k0], 4 K[k]}, {5 K[k0], 5 K[k]}, {6 K[k0], 6 K[k]}}, {-10, -5, 5, 10}};
(c E1 - s d)/c /. fromcsdErule1;
plot1 = Plot[(% /. {k -> k0}), {p, 0, 6 K[k0]}, PlotStyle -> {Red, Dashed, Thick}, Ticks -> tic, PlotLegends -> {Subscript[g, 1][p]}];
plot2 = Plot[(%% /. {k -> k0}), {p, 0, 6 K[k0]}, PlotStyle -> {Thick}, Ticks -> tic, PlotLegends -> {Subscript[f, 1][p]}];
plot3 = Show[plot1, plot2, AxesLabel -> {p, Subscript[f, 1][p] Subscript[g, 1][p]}, TicksStyle -> Directive[12], LabelStyle -> Directive[Bold, 14], ImageSize -> Large]
Export["f1g1.png", plot3]
```

B.5 rules.txt

Note that in this file 's' represents $\mathrm{sn}(p, k)$ and 'ts' represents $\mathrm{sn}(\tau, k)$. Same convetion applies to functions cn, dn etc.

```
ctosrules := {c^2 -> 1 - s^2, c^3 -> c (1 - s^2), c^4 -> (1 - s^2)^2, c^5 -> c (1 - s^2)^2, c^6 -> (1 - s^2)^3, c^7 -> c (1 - s^2)^3, c^8 -> (1 - s^2)^4, c^9 -> c (1 - s^2)^4, c^10 -> (1 - s^2)^5, c^11 -> c (1 - s^2)^5, c^12 -> (1 - s^2)^6, c^13 -> c (1 - s^2)^6, c^14 -> (1 - s^2)^7, c^15 -> c (1 - s^2)^7, c^16 -> (1 - s^2)^8, c^17 -> c (1 - s^2)^8, c^18 -> (1 - s^2)^9, c^19 -> c (1 - s^2)^9, c^20 -> (1 - s^2)^10, c^21 -> c (1 - s^2)^10, c^22 -> (1 - s^2)^11, c^23 -> c (1 - s^2)^11, c^24 -> (1 - s^2)^12, c^25 -> c (1 - s^2)^12, c^26 -> (1 - s^2)^13, c^27 -> c (1 - s^2)^13, c^28 -> (1 - s^2)^14, c^29 -> c (1 - s^2)^14, c^30 -> (1 - s^2)^15}

stocrules := {s^2 -> 1 - c^2, s^3 -> s (1 - c^2), s^4 -> (1 - c^2)^2, s^5 -> s (1 - c^2)^2, s^6 -> (1 - c^2)^3, s^7 -> s (1 - c^2)^3, s^8 -> (1 - c^2)^4, s^9 -> s (1 - c^2)^4, s^10 -> (1 - c^2)^5, s^11 -> s (1 - c^2)^5, s^12 -> (1 - c^2)^6, s^13 -> s (1 - c^2)^6, s^14 -> (1 - c^2)^7, s^15 -> s (1 - c^2)^7, s^16 -> (1 - c^2)^8, s^17 -> s (1 - c^2)^8, s^18 -> (1 - c^2)^9, s^19 -> s (1 - c^2)^9, s^20 -> (1 - c^2)^10, s^21 -> s (1 - c^2)^10, s^22 -> (1 - c^2)^11, s^23 -> s (1 - c^2)^11, s^24 -> (1 - c^2)^12, s^25 -> s (1 - c^2)^12, s^26 -> (1 - c^2)^13, s^27 -> s (1 - c^2)^13, s^28 -> (1 - c^2)^14, s^29 -> s (1 - c^2)^14, s^30 -> (1 - c^2)^15}

stodrules := {s^2 -> (1 - d^2)/k^2, s^3 -> s (1 - d^2)/k^2, s^4 -> ((1 - d^2)/k^2)^2, s^5 -> s ((1 - d^2)/k^2)^2, s^6 -> ((1 - d^2)/k^2)^3, s^7 -> s ((1 - d^2)/k^2)^3, s^8 -> ((1 - d^2)/k^2)^4, s^9 -> s ((1 - d^2)/k^2)^4, s^10 -> ((1 - d^2)/k^2)^5, s^11 -> s ((1 - d^2)/k^2)^5, s^12 -> ((1 - d^2)/k^2)^6, s^13 -> s ((1 - d^2)/k^2)^6, s^14 -> ((1 - d^2)/k^2)^7, s^15 -> s ((1 - d^2)/k^2)^7, s^16 -> ((1 - d^2)/k^2)^8, s^17 -> s ((1
```

- d^2)/k^2)^8, s^18 -> ((1 - d^2)/k^2)^9, s^19 -> s ((1 - d^2)/k^2)^9, s^20 -> ((1 - d^2)/k^2)^10, s^21 -> s ((1 - d^2)/k^2)^10, s^22 -> ((1 - d^2)/k^2)^11, s^23 -> s ((1 - d^2)/k^2)^11, s^24 -> ((1 - d^2)/k^2)^12, s^25 -> s ((1 - d^2)/k^2)^12, s^26 -> ((1 - d^2)/k^2)^13, s^27 -> s ((1 - d^2)/k^2)^13, s^28 -> ((1 - d^2)/k^2)^14, s^29 -> s ((1 - d^2)/k^2)^14, s^30 -> ((1 - d^2)/k^2)^15, 1-k^2 s^4-> (c^2+d^2 s^2)}

dtosrules := {d^2 -> 1 - k^2 s^2, d^3 -> d (1 - k^2 s^2), d^4 -> (1 - k^2 s^2)^2, d^5 -> d (1 - k^2 s^2)^2, d^6 -> (1 - k^2 s^2)^3, d^7 -> d (1 - k^2 s^2)^3, d^8 -> (1 - k^2 s^2)^4, d^9 -> d (1 - k^2 s^2)^4, d^10 -> (1 - k^2 s^2)^5, d^11 -> d (1 - k^2 s^2)^5, d^12 -> (1 - k^2 s^2)^6, d^13 -> d (1 - k^2 s^2)^6, d^14 -> (1 - k^2 s^2)^7, d^15 -> d (1 - k^2 s^2)^7, d^16 -> (1 - k^2 s^2)^8, d^17 -> d (1 - k^2 s^2)^8, d^18 -> (1 - k^2 s^2)^9, d^19 -> d (1 - k^2 s^2)^9, d^20 -> (1 - k^2 s^2)^10, d^21 -> d (1 - k^2 s^2)^10, d^22 -> (1 - k^2 s^2)^11, d^23 -> d (1 - k^2 s^2)^11, d^24 -> (1 - k^2 s^2)^12, d^25 -> d (1 - k^2 s^2)^12, d^26 -> (1 - k^2 s^2)^13, d^27 -> d (1 - k^2 s^2)^13, d^28 -> (1 - k^2 s^2)^14, d^29 -> d (1 - k^2 s^2)^14, d^30 -> (1 - k^2 s^2)^15}

ctosrules1 := {c^2 -> 1 - s^2, c^3 -> c (1 - s^2), c^4 -> (1 - s^2)^2, c^5 -> c (1 - s^2)^2, c^6 -> (1 - s^2)^3, c^7 -> c (1 - s^2)^3, c^8 -> (1 - s^2)^4, c^9 -> c (1 - s^2)^4, c^10 -> (1 - s^2)^5, c^11 -> c (1 - s^2)^5, c^12 -> (1 - s^2)^6, c^13 -> c (1 - s^2)^6, c^14 -> (1 - s^2)^7, c^15 -> c (1 - s^2)^7, c^16 -> (1 - s^2)^8, c^17 -> c (1 - s^2)^8, c^18 -> (1 - s^2)^9, c^19 -> c (1 - s^2)^9, c^20 -> (1 - s^2)^10, c^21 -> c (1 - s^2)^10, c^22 -> (1 - s^2)^11, c^23 -> c (1 - s^2)^11, c^24 -> (1 - s^2)^12, c^25 -> c (1 - s^2)^12, c^26 -> (1 - s^2)^13, c^27 -> c (1 - s^2)^13, c^28 -> (1 - s^2)^14, c^29 -> c (1 - s^2)^14, c^30 -> (1 - s^2)^15, tc^2 -> 1 - ts^2, tc^3 -> tc (1 - ts^2), tc^4 -> (1 - ts^2)^2, tc^5 -> tc (1 - ts^2)^2, tc^6 -> (1 - ts^2)^3, tc^7 -> tc (1 - ts^2)^3, tc^8 -> (1 - ts^2)^4, tc^9 -> tc (1 - ts^2)^4, tc^10 -> (1 - ts^2)^5, tc^11 -> tc (1 - ts^2)^5, tc^12 -> (1 - ts^2)^6, tc^13 -> tc (1 - ts^2)^6, tc^14 -> (1 - ts^2)^7, tc^15 -> tc (1 - ts^2)^7, tc^16 -> (1 - ts^2)^8, tc^17 -> tc (1 - ts^2)^8, tc^18 -> (1 - ts^2)^9, tc^19 -> tc (1 - ts^2)^9, tc^20 -> (1 - ts^2)^10, tc^21 -> tc (1 - ts^2)^10, tc^22 -> (1 - ts^2)^11, tc^23 -> tc (1 - ts^2)^11, tc^24 -> (1 - ts^2)^12, tc^25 -> tc (1 - ts^2)^12, tc^26 -> (1 - ts^2)^13, tc^27 -> tc (1 - ts^2)^13, tc^28 -> (1 - ts^2)^14, tc^29 -> tc (1 - ts^2)^14, tc^30 -> (1 - ts^2)^15}

stocrules1 := {s^2 -> 1 - c^2, s^3 -> s (1 - c^2), s^4 -> (1 - c^2)^2, s^5 -> s (1 - c^2)^2, s^6 -> (1 - c^2)^3, s^7 -> s (1 - c^2)^3, s^8 -> (1 - c^2)^4, s^9 -> s (1 - c^2)^4, s^10 -> (1 - c^2)^5, s^11 -> s (1 - c^2)^5, s^12 -> (1 - c^2)^6, s^13 -> s (1 - c^2)^6, s^14 -> (1 - c^2)^7, s^15 -> s (1 - c^2)^7, s^16 -> (1 - c^2)^8, s^17 -> s (1 - c^2)^8, s^18 -> (1 - c^2)^9, s^19 -> s (1 - c^2)^9, s^20 -> (1 - c^2)^10, s^21 -> s (1 - c^2)^10, s^22 -> (1 - c^2)^11, s^23 -> s (1 - c^2)^11, s^24 -> (1 - c^2)^12, s^25 -> s (1 - c^2)^12, s^26 -> (1 - c^2)^13, s^27 -> s (1 - c^2)^13, s^28 -> (1 - c^2)^14, s^29 -> s (1 - c^2)^14, s^30 -> (1 - c^2)^15, ts^2 -> 1 - tc^2, ts^3 -> ts (1 - tc^2), ts^4 -> (1 - tc^2)^2, ts^5 -> ts (1 - tc^2)^2, ts^6 -> (1 - tc^2)^3, ts^7 -> ts (1 - tc^2)^3, ts^8 -> (1 - tc^2)^4, ts^9 -> ts (1 -

tc^2)^4, ts^10 -> (1 - tc^2)^5, ts^11 -> ts (1 - tc^2)^5, ts^12 -> (1 - tc^2)^6, ts^13 -> ts (1 - tc^2)^6, ts^14 -> (1 - tc^2)^7, ts^15 -> ts (1 - tc^2)^7, ts^16 -> (1 - tc^2)^8, ts^17 -> ts (1 - tc^2)^8, ts^18 -> (1 - tc^2)^9, ts^19 -> ts (1 - tc^2)^9, ts^20 -> (1 - tc^2)^10, ts^21 -> ts (1 - tc^2)^10, ts^22 -> (1 - tc^2)^11, ts^23 -> ts (1 - tc^2)^11, ts^24 -> (1 - tc^2)^12, ts^25 -> ts (1 - tc^2)^12, ts^26 -> (1 - tc^2)^13, ts^27 -> ts (1 - tc^2)^13, ts^28 -> (1 - tc^2)^14, ts^29 -> ts (1 - tc^2)^14, ts^30 -> (1 - tc^2)^15}

stodrules1 := {s^2 -> (1 - d^2)/k^2, s^3 -> s (1 - d^2)/k^2, s^4 -> ((1 - d^2)/k^2)^2, s^5 -> s ((1 - d^2)/k^2)^2, s^6 -> ((1 - d^2)/k^2)^3, s^7 -> s ((1 - d^2)/k^2)^3, s^8 -> ((1 - d^2)/k^2)^4, s^9 -> s ((1 - d^2)/k^2)^4, s^10 -> ((1 - d^2)/k^2)^5, s^11 -> s ((1 - d^2)/k^2)^5, s^12 -> ((1 - d^2)/k^2)^6, s^13 -> s ((1 - d^2)/k^2)^6, s^14 -> ((1 - d^2)/k^2)^7, s^15 -> s ((1 - d^2)/k^2)^7, s^16 -> ((1 - d^2)/k^2)^8, s^17 -> s ((1 - d^2)/k^2)^8, s^18 -> ((1 - d^2)/k^2)^9, s^19 -> s ((1 - d^2)/k^2)^9, s^20 -> ((1 - d^2)/k^2)^10, s^21 -> s ((1 - d^2)/k^2)^10, s^22 -> ((1 - d^2)/k^2)^11, s^23 -> s ((1 - d^2)/k^2)^11, s^24 -> ((1 - d^2)/k^2)^12, s^25 -> s ((1 - d^2)/k^2)^12, s^26 -> ((1 - d^2)/k^2)^13, s^27 -> s ((1 - d^2)/k^2)^13, s^28 -> ((1 - d^2)/k^2)^14, s^29 -> s ((1 - d^2)/k^2)^14, s^30 -> ((1 - d^2)/k^2)^15, ts^2 -> (1 - td^2)/k^2, ts^3 -> ts (1 - td^2)/k^2, ts^4 -> ((1 - td^2)/k^2)^2, ts^5 -> ts ((1 - td^2)/k^2)^2, ts^6 -> ((1 - td^2)/k^2)^3, ts^7 -> ts ((1 - td^2)/k^2)^3, ts^8 -> ((1 - td^2)/k^2)^4, ts^9 -> ts ((1 - td^2)/k^2)^4, ts^10 -> ((1 - td^2)/k^2)^5, ts^11 -> ts ((1 - td^2)/k^2)^5, ts^12 -> ((1 - td^2)/k^2)^6, ts^13 -> ts ((1 - td^2)/k^2)^6, ts^14 -> ((1 - td^2)/k^2)^7, ts^15 -> ts ((1 - td^2)/k^2)^7, ts^16 -> ((1 - td^2)/k^2)^8, ts^17 -> ts ((1 - td^2)/k^2)^8, ts^18 -> ((1 - td^2)/k^2)^9, ts^19 -> ts ((1 - td^2)/k^2)^9, ts^20 -> ((1 - td^2)/k^2)^10, ts^21 -> ts ((1 - td^2)/k^2)^10, ts^22 -> ((1 - td^2)/k^2)^11, ts^23 -> ts ((1 - td^2)/k^2)^11, ts^24 -> ((1 - td^2)/k^2)^12, ts^25 -> ts ((1 - td^2)/k^2)^12, ts^26 -> ((1 - td^2)/k^2)^13, ts^27 -> ts ((1 - td^2)/k^2)^13, ts^28 -> ((1 - td^2)/k^2)^14, ts^29 -> ts ((1 - td^2)/k^2)^14, ts^30 -> ((1 - td^2)/k^2)^15}

dtosrules1 := {d^2 -> 1 - k^2 s^2, d^3 -> d (1 - k^2 s^2), d^4 -> (1 - k^2 s^2)^2, d^5 -> d (1 - k^2 s^2)^2, d^6 -> (1 - k^2 s^2)^3, d^7 -> d (1 - k^2 s^2)^3, d^8 -> (1 - k^2 s^2)^4, d^9 -> d (1 - k^2 s^2)^4, d^10 -> (1 - k^2 s^2)^5, d^11 -> d (1 - k^2 s^2)^5, d^12 -> (1 - k^2 s^2)^6, d^13 -> d (1 - k^2 s^2)^6, d^14 -> (1 - k^2 s^2)^7, d^15 -> d (1 - k^2 s^2)^7, d^16 -> (1 - k^2 s^2)^8, d^17 -> d (1 - k^2 s^2)^8, d^18 -> (1 - k^2 s^2)^9, d^19 -> d (1 - k^2 s^2)^9, d^20 -> (1 - k^2 s^2)^10, d^21 -> d (1 - k^2 s^2)^10, d^22 -> (1 - k^2 s^2)^11, d^23 -> d (1 - k^2 s^2)^11, d^24 -> (1 - k^2 s^2)^12, d^25 -> d (1 - k^2 s^2)^12, d^26 -> (1 - k^2 s^2)^13, d^27 -> d (1 - k^2 s^2)^13, d^28 -> (1 - k^2 s^2)^14, d^29 -> d (1 - k^2 s^2)^14, d^30 -> (1 - k^2 s^2)^15, td^2 -> 1 - k^2 ts^2, td^3 -> td (1 - k^2 ts^2), td^4 -> (1 - k^2 ts^2)^2, td^5 -> td (1 - k^2 ts^2)^2, td^6 -> (1 - k^2 ts^2)^3, td^7 -> td (1 - k^2 ts^2)^3, td^8 -> (1 - k^2 ts^2)^4, td^9 -> td (1 - k^2 ts^2)^4, td^10 -> (1 - k^2 ts^2)^5, td^11 -> td (1 - k^2 ts^2)^5, td^12 -> (1

- k^2 ts^2)^6, td^13 -> td (1 - k^2 ts^2)^6, td^14 -> (1 - k^2 ts^2)^7, td^15 -> td (1 - k^2 ts^2)^7, td^16 -> (1 - k^2 ts^2)^8, td^17 -> td (1 - k^2 ts^2)^8, td^18 -> (1 - k^2 ts^2)^9, td^19 -> td (1 - k^2 ts^2)^9, td^20 -> (1 - k^2 ts^2)^10, td^21 -> td (1 - k^2 ts^2)^10, td^22 -> (1 - k^2 ts^2)^11, td^23 -> td (1 - k^2 ts^2)^11, td^24 -> (1 - k^2 ts^2)^12, td^25 -> td (1 - k^2 ts^2)^12, td^26 -> (1 - k^2 ts^2)^13, td^27 -> td (1 - k^2 ts^2)^13, td^28 -> (1 - k^2 ts^2)^14, td^29 -> td (1 - k^2 ts^2)^14, td^30 -> (1 - k^2 ts^2)^15}

tocsdErule1 := {cn[p, k] -> c, sn[p, k] -> s, dn[p, k] -> d, EE[p, k] -> E1, K[k] -> K, cn[tau, k] -> tc, sn[tau, k] -> ts, dn[tau, k] -> td, EE[tau, k] -> tE1} fromcsdErule1 := {c -> cn[p, k], s -> sn[p, k], d -> dn[p, k], K -> K[k], E1 -> EE[p, k], E2 -> EEE[k], tc -> cn[tau, k], ts -> sn[tau, k], td -> dn[tau, k], tE1 -> EE[tau, k]}

Appendix C

Mathematica Code for Calculation with the Jacobian of Exponential Mapping $\lambda \in C_1$

From our discussion of the conjugate loci in chapter 8 it is clear that the conjugate loci exist at the roots of Jacobian J of the exponential mapping. Following Mathematica code calculates and simplifies the expression of Jacobian in terms of variables p and τ. General discussion and scheme of simplification is same as discussed in Appendix B. Here again we are interesting in finding the first root of Jacobian J in terms of time which will give the first Maxwell time t_1^{MAX}.

The code is divided into five sections. In Section C.1 we give the setup code including the definition of addition formulas for elliptic functions. We also give formulas for the extremal trajectories (x_t, y_t, z_t). In section 2 we calculate and simplify the expression of Jacobian. In section 3 we validate the results of simplification by three methods. First we plot the original expression of Jacobian for some value of k, τ and p and compare the results. Then we take the difference of original expression of Jacobian and final simplified expression. After applying same set of transformations, the result is 0 as expected. Finally we also plot the error between original and simplified expression of Jacobian. It turns out that the expressions agree upto the numerical error. In section 4 we give code for obtaining Taylor expansion of the Jacobian. In section 5 we calculate the roots of Jacobian numerically and plot them along side the upper and lower bounds of the first conjugate time. The plots show that the first conjugate time for $k \in (0, 1)$ lie within the bounds that were proved analytically in chapter 8. We have also included the output of validation process for the interested reader. The pseduo code for this corresponding Mathematica code is given as:

1. Initialize notebook directory

2. Define the addition / subtraction formulas for Jacobi elliptic functions

3. Define transformation rules / identities for Jacobi elliptic functions
4. Define parametric equations of extremal trajectories
5. Compute the expression for Jacobian J
6. Simplify the Jacobian expression using builtin Mathematica functions and transformation rules
7. Obtain the roots of $J = 0$ from simplified expression
8. Plot the original and simplified function for comparison

C.1 Setup

```
(*Elliptic Functions*)
amm[phi_, k_] := JacobiAmplitude[phi, k^2]
EE[phi_, k_] := EllipticE[amm[phi, k], k^2]
K[k_] := EllipticK[k^2]
EEE[k_] := EllipticE[k^2]
sn[p_, k_] := JacobiSN[p, k^2]
cn[p_, k_] := JacobiCN[p, k^2]
dn[p_, k_] := JacobiDN[p, k^2]
(*Addition formulas*) (*We first unprotect a function to modify its definition. In the end \
we again protect*)
Unprotect[JacobiDN]
(*dn(p+tau)*)
JacobiDN[p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] - k^2*JacobiSN[p, k^2]*JacobiCN[p,
k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*dn(-p+tau)*)
JacobiDN[-p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] + k^2*JacobiSN[p,
k^2]*JacobiCN[p, k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[t
k^2]^2)
Protect[JacobiDN]

Unprotect[EllipticE]
(*E(p+tau)*)
EllipticE[JacobiAmplitude[p + tau, k^2], k^2] := EllipticE[JacobiAmplitude[p, k^2], k^2] +
EllipticE[JacobiAmplitude[tau, k^2], k^2] - k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[p
+ tau, k^2]
(*E(-p+tau)*)
```

```
EllipticE[JacobiAmplitude[-p + tau, k^2], k^2] := -EllipticE[JacobiAmplitude[p, k^2], k^2]
+ EllipticE[JacobiAmplitude[tau, k^2], k^2] + k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[-
p + tau, k^2]
```

Protect[EllipticE]

Unprotect[JacobiSN]

```
(*sn(p+tau)*)
JacobiSN[p + tau, k^2] := (JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] + Ja-
cobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*sn(-p+tau)*)
JacobiSN[-p + tau, k^2] := (-JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] +
JacobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
```

Protect[JacobiSN]

Unprotect[JacobiCN]

```
(*cn(p+tau)*)
JacobiCN[p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] - JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*cn(-p+tau)*)
JacobiCN[-p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] + JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
```

Protect[JacobiCN]

```
(*Definitions of (x, y, z) coordinates of Extremal Trajectories Function*)
phit[phhi0_, t_] := t + phhi0;
w[phhi0_, k_] := 1/(dn[phhi0, k] - k cn[phhi0, k]);
x11[phhi0_, k_] := w[phhi0, k] + 1/(w[phhi0, k] (1 - k^2));
x12[phhi0_, k_] := k/(w[phhi0, k] (1 - k^2)) - k w[phhi0, k];
y11[phhi0_, k_] := w[phhi0, k] - 1/(w[phhi0, k] (1 - k^2));
y12[phhi0_, k_] := k w[phhi0, k] + k/(w[phhi0, k] (1 - k^2));
x[t_, phhi0_, k_] := 1/2 (x11[phhi0, k] (EE[phit[phhi0, t], k] - EE[phhi0, k]) + x12[phhi0,
k] (sn[phit[phhi0, t], k] - sn[phhi0, k]));
y[t_, phhi0_, k_] := 1/2 (y11[phhi0, k] (EE[phit[phhi0, t], k] - EE[phhi0, k]) - y12[phhi0,
k] (sn[phit[phhi0, t], k] - sn[phhi0, k]));
z[t_, phhi0_, k_] := Log[w[phhi0, k] (dn[phit[phhi0, t], k] - k cn[phit[phhi0, t], k])];
ze[t_, phhi0_, k_] := Exp[z[t, phhi0, k]];
```

C.2 Jacobian and Simplification

```
(* We calculate the Jacobian using exp(z) instead of z b/c z is an \ expression involving natural log and since we are looking for roots \ of the Jacobian and not the Jacobian itself we should be doing fine. *)
det = Det[ D[{x[t, phhi0, k], y[t, phhi0, k], ze[t, phhi0, k]}, {{t, phhi0, k}}]];        (*Calculation of Jacobian*)

% /. {JacobiSC[phhi0, k^2] -> JacobiSN[phhi0, k^2]/JacobiCN[phhi0, k^2]} /. {JacobiSC[ phhi0 + t, k^2] -> JacobiSN[phhi0 + t, k^2]/JacobiCN[phhi0 + t, k^2]} /. {JacobiCD[ phhi0, k^2] -> JacobiCN[phhi0, k^2]/JacobiDN[phhi0, k^2]} /. {JacobiCD[ phhi0 + t, k^2] -> JacobiCN[phhi0 + t, k^2]/JacobiDN[phhi0 + t, k^2]};
% /. {phhi0 -> tau - p} /. {t -> 2 p};        (*Introduce change of coordinates*)
% /. {EllipticF[JacobiAmplitude[p - tau, k^2], k^2] -> p - tau} /. {EllipticF[JacobiAmplitude[p + tau, k^2], k^2] -> p + tau};
% /. tocsdErule1;
Together[%];
d1 = Denominator[%];
x1 = Numerator[%%];
Factor[x1];
Together[%] /. dtosrules1 /. ctosrules1;
x2 = Simplify[%];        (*First stage of simplification*)

(*Following expression appears numerous times. It is simplified manually and replaced in the expression of Jacobian. The manual simplification is appended after the code.*)
x3 = x2 /. {Sqrt[ 1 - (k^2*(s*tc*td - c*d*ts)^2)/(-1 + k^2*s^2*ts^2)^2] -> -(d td + c k^2 s tc ts)/(-1 + k^2 s^2 ts^2)} /. {Sqrt[ 1 - (k^2*(s*tc*td + c*d*ts)^2)/(-1 + k^2*s^2*ts^2)^2] -> -(d td - c k^2 s tc ts)/(-1 + k^2 s^2 ts^2)};
Together[x3] /. dtosrules1 /. ctosrules1;
x4 = FullSimplify[%];
(* Following expression is obtained by manual simplification which is documented separately. The actual expression is x4=(-4 k (-1+k^2 s^2 ts^2)^5 )(2 d E1 k s tc-d k p s tc+d k^3 p s \ tc-c k s^2 tc+c E1^2 k s^2 tc-c k^3 s^2 tc-c E1 k p s^2 tc+c E1 k^3 p \ s^2 tc-2 d E1 k s^3 tc+d k p s^3 tc-d k^3 p s^3 tc+c k^3 s^4 tc-2 c \ E1 s td+c p s td-c k^2 p s td+d s^2 td-d E1^2 s^2 td+d k^2 s^2 td+d \ E1 p s^2 td-d E1 k^2 p s^2 td+2 c E1 k^2 s^3 td-c k^2 p s^3 td+c k^4 \ p s^3 td-d k^2 s^4 td-c E1^2 k tc ts^2+c E1 k p tc ts^2-c E1 k^3 p tc \ ts^2+c k^3 s^2 tc ts^2+d E1^2 td ts^2-d E1 p td ts^2+d E1 k^2 p td \ ts^2-d k^2 s^2 td ts^2) *)

x5 = (c k tc - d td)*(s c d*(2 E1 - p + k^2 p) - d^2 s^2 - k^2 s^2 tc ^2 + E1*(s^2 - ts^2)*(E1 - p + k^2 p))*(-4 k (-1 + k^2 s^2 ts^2)^5 );        (*Final simplified expression of
```

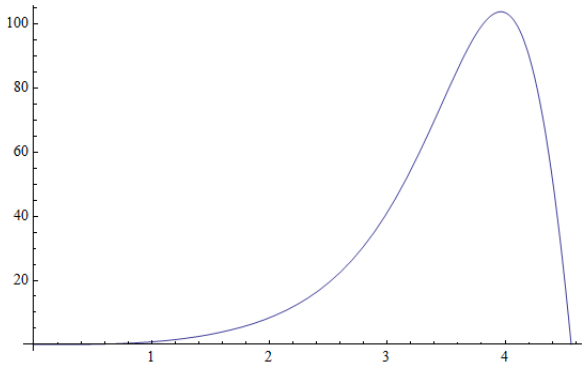

Figure C.1: Original Jacobian for $k = 0.9$, $\tau = 0$ and $0 \leq p \leq 2K[k]$

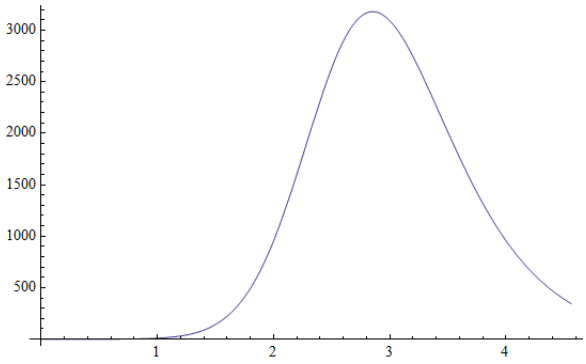

Figure C.2: Original Jacobian for $k = 0.9$, $\tau = K[k]$ and $0 \leq p \leq 2K[k]$

Jacobian*)

C.3 Validation of Simplified Expression of Jacobian

```
Jnew1 = (x2/d1) /. fromcsdErule1 // Simplify; (*Plot of original expression of the Jacobian*)
k0 := 0.9;
(Jnew1 /. {k -> k0, tau -> 0});
plot11new = Plot[{%}, {p, 0, 2 K[k0]}]
(Jnew1 /. {k -> k0, tau -> K[k0]});
plot12new = Plot[{%}, {p, 0, 2 K[k0]}]

Jnew5 = (x5/d1) /. fromcsdErule1 // Simplify; (*Plot of simplified expression of the Jacobian*)
k0 := 0.9;
(Jnew5 /. {k -> k0, tau -> 0});
plot51new = Plot[{%}, {p, 0, 2 K[k0]}]
(Jnew5 /. {k -> k0, tau -> K[k0]});
plot52new = Plot[{%}, {p, 0, 2 K[k0]}]
```

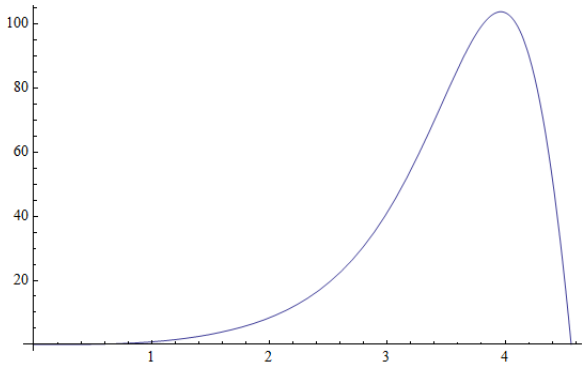

Figure C.3: Simplified Jacobian for $k = 0.9$, $\tau = 0$ and $0 \leq p \leq 2K[k]$

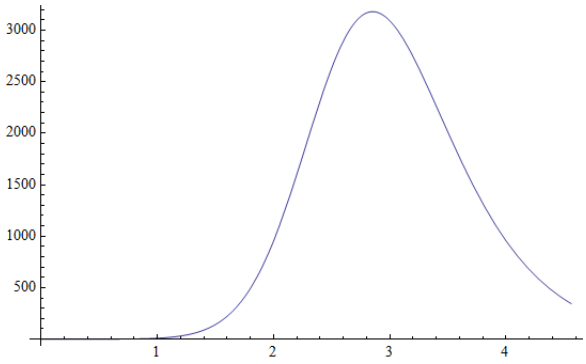

Figure C.4: Simplified Jacobian for $k = 0.9$, $\tau = K[k]$ and $0 \leq p \leq 2K[k]$

```
(* Validation of simplification by taking difference of original expression of Jacobian and
simplified expression of Jacobian
Jnew5-Det[D[{x[t, phhi0, k], y[t, phhi0, k], ze[t, phhi0, k]}, {{t, phhi0, k}}]]*)
Jnew5 - det;
% /. {JacobiSC[phhi0, k^2] -> JacobiSN[phhi0, k^2]/JacobiCN[phhi0, k^2]} /. {JacobiSC[
phhi0 + t, k^2] -> JacobiSN[phhi0 + t, k^2]/JacobiCN[phhi0 + t, k^2]} /. {JacobiCD[
phhi0, k^2] -> JacobiCN[phhi0, k^2]/JacobiDN[phhi0, k^2]} /. {JacobiCD[ phhi0 + t, k^2]
-> JacobiCN[phhi0 + t, k^2]/JacobiDN[phhi0 + t, k^2]};
% /. {phhi0 -> tau - p} /. {t -> 2 p};
% /. {EllipticF[JacobiAmplitude[p - tau, k^2], k^2] -> p - tau} /. {EllipticF[JacobiAmplitude[p
+ tau, k^2], k^2] -> p + tau};
% /. tocsdErule1;
Together[%];
d1s = Denominator[%];
x1s = Numerator[%%];
Factor[x1s];
Together[%] /. dtosrules1 /. ctosrules1;
x2s = Simplify[%];
```

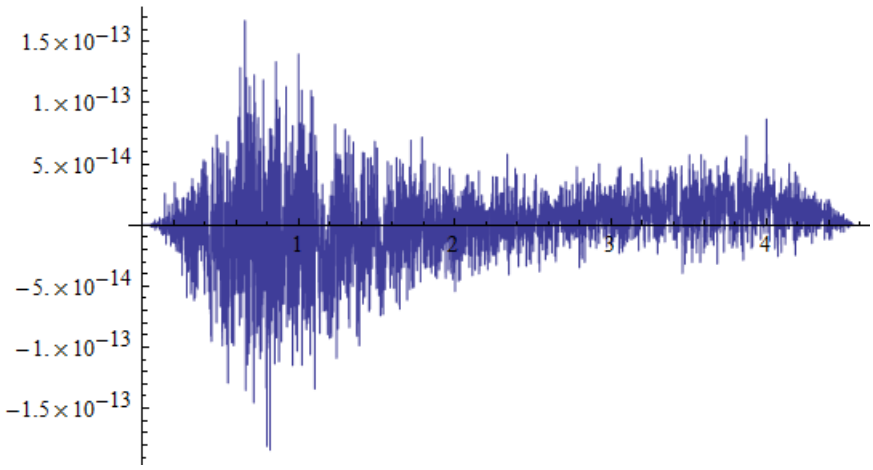

Figure C.5: Difference between Original and Simplified Jacobian for $k = 0.9$, $\tau = 0$ and $0 \leq p \leq 2K[k]$

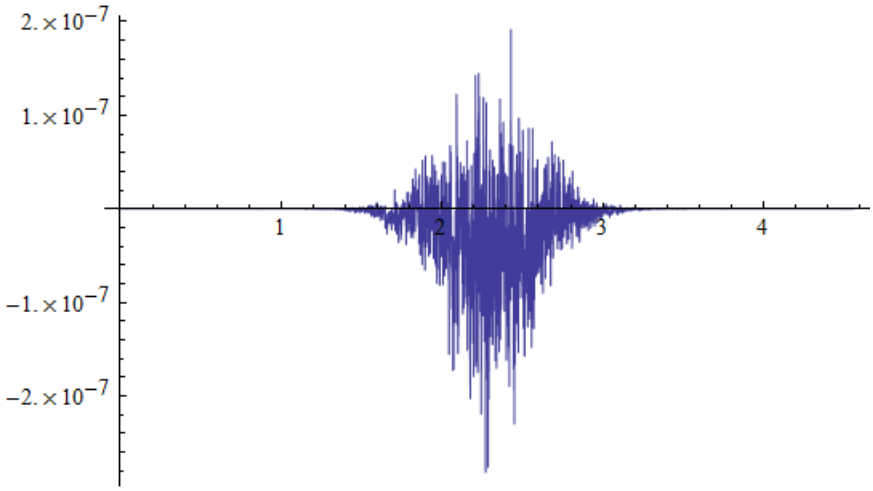

Figure C.6: Difference between Original and Simplified Jacobian for $k = 0.9$, $\tau = K(k)$ and $0 \leq p \leq 2K(k)$

```
x3s = x2s /. {Sqrt[ 1 - (k^2*(s*tc*td - c*d*ts)^2)/(-1 + k^2*s^2*ts^2)^2] -> -(d td + c k^2 s
tc ts)/(-1 + k^2 s^2 ts^2)} /. {Sqrt[ 1 - (k^2*(s*tc*td + c*d*ts)^2)/(-1 + k^2*s^2*ts^2)^2]
-> -(d td - c k^2 s tc ts)/(-1 + k^2 s^2 ts^2)};
Together[x3s] /. dtosrules1 /. ctosrules1;
x4s = FullSimplify[%]      (*The result of last command is 0*)

(*Validation by error analysis*)
Jnew6 = ((x5 - x1)/d1) /. fromcsdErule1 // Simplify;
k0 := 0.9;
(Jnew6 /. {k -> k0, tau -> 0});
plot61new = Plot[{%}, {p, 0, 2 K[k0]}, PlotRange -> All]
(Jnew6 /. {k -> k0, tau -> K[k0]});
plot62new = Plot[{%}, {p, 0, 2 K[k0]}, PlotRange -> All]
```

C.4 Taylor Expansion of Jacobian for SH(2)

```
(*(k^2+p^2)->0*)
x5 = (c k tc - d td)*(s c d*(2 E1 - p + k^2 p) - d^2 s^2 - k^2 s^2 tc ^2 + E1*(s^2 -
ts^2)*(E1 - p + k^2 p))*(-4 k (-1 + k^2 s^2 ts^2)^5 );
d1 = (-1 + k^2)^2 (c k tc - d td) (-1 + k s ts)^7 (1 + k s ts)^5;
v3 = (x5/(d1)) /. fromcsdErule1 // Simplify;
Series[v3, {p, 0, 4}] // FullSimplify
Series[v3, {k, 0, 1}] // FullSimplify
Series[v3, {k, 0, 1}, {p, 0, 4}] // Normal

(*k->0*)
Series[v3, {k, 0, 1}];
% /. tocsdErule1;
Normal[%];
FullSimplify[%]
```

C.5 Numerical Calculation of Roots of Jacobian

```
ki = Range[0.01, 0.99, 0.05];
pi = 0*ki;
ub = 0*ki;
f1[p_, k_] := EE[p, k]*cn[p, k] - sn[p, k]*dn[p, k];
nm1 = Numerator[v3];
For[i = 1, i <= Length[ki], i++, k0 = ki[[i]];
nm2 = nm1 /. tau -> K[k0] /. k -> k0;
fi = f1[p, k] /. k -> k0;
pi[[i]] = p /. FindRoot[nm2 == 0, {p, 4 K[k0]}];
ub[[i]] = p /. FindRoot[fi == 0, {p, 4.5 K[k0]}]; ]
Partition[Riffle[ki, pi], 2]
plot1 = ListPlot[%, PlotRange -> {0, 10}];
Partition[Riffle[ki, ub], 2]
plot11 = ListPlot[%, PlotRange -> {0, 10}, PlotStyle -> RGBColor[1, 0, 0]];
lb = 4 K[ki]; Partition[Riffle[ki, lb], 2]
plot111 = ListPlot[%, PlotRange -> {0, 10}, PlotStyle -> RGBColor[0, 1, 0]];
Show[plot1, plot11, plot111, PlotRange -> {0, 10}, BaselinePosition -> Bottom]
```

Appendix D

Mathematica Code for Plotting 3-Dimensional Sub-Riemannian Objects

In chapter 8 we presented the plots of sub-Riemannian sphere, sub-Riemannian wavefront and Matryoshka of sub-Riemannian wavefront. In the following we present the Mathematica code along with the transformation that was used to produce the plots.

From our discussion of the conjugate loci in chapter 8 it is clear that the conjugate loci exist at the roots of Jacobian J of the exponential mapping. Following Mathematica code calculates and simplifies the expression of Jacobian in terms of variables p and τ. General discussion and scheme of simplification is same as discussed in Appendix B. Here again we are interesting in finding the first root of Jacobian J in terms of time which will give the first Maxwell time t_1^{MAX}.

The code is divided into five sections. In Section D.1 we give the setup code including the definition of addition formulas for elliptic functions. We also give formulas for the extremal trajectories (x_t, y_t, z_t). In section 2 we calculate and simplify the expression of Jacobian. In section 3 we validate the results of simplification by three methods. First we plot the original expression of Jacobian for some value of k, τ and p and compare the results. Then we take the difference of original expression of Jacobian and final simplified expression. After applying same set of transformations, the result is 0 as expected. Finally we also plot the error between original and simplified expression of Jacobian. It turns out that the expressions agree upto the numerical error. In section 4 we give code for obtaining Taylor expansion of the Jacobian. In section 5 we calculate the roots of Jacobian numerically and plot them along side the upper and lower bounds of the first conjugate time. The plots show that the first conjugate time for $k \in (0, 1)$ lie within the bounds that were proved analytically in chapter 8. We have also included the output of validation process for the interested reader. The pseduo code for this corresponding Mathematica code is given as:

1. Initialize notebook directory

2. Define the addition / subtraction formulas for Jacobi elliptic functions

3. Define transformation rules / identities for Jacobi elliptic functions
4. Define parametric equations of extremal trajectories and the function $(R_1(q), R_2(q), z)$
5. Transform the equations in terms of trigonometric functions for speedy potting
6. Obtain 3D parametric plots of the functions in terms of rectifying coordinates

D.1 Plot of Part of the Wavefront for Time t = R Corresponding to $\lambda \in C_1$

Setup

```
(*Elliptic Functions*)
amm[phi_, k_] := JacobiAmplitude[phi, k^2]
EE[phi_, k_] := EllipticE[amm[phi, k], k^2]
K[k_] := EllipticK[k^2]
EEE[k_] := EllipticE[k^2]
sn[p_, k_] := JacobiSN[p, k^2]
cn[p_, k_] := JacobiCN[p, k^2]
dn[p_, k_] := JacobiDN[p, k^2]
(*Addition formulas*) (*We first unprotect a function to modify its definition. In the end \
we again protect*)
Unprotect[JacobiDN]
(*dn(p+tau)*)
JacobiDN[p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] - k^2*JacobiSN[p, k^2]*JacobiCN[p,
k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*dn(-p+tau)*)
JacobiDN[-p + tau, k^2] := (JacobiDN[p, k^2]*JacobiDN[tau, k^2] + k^2*JacobiSN[p,
k^2]*JacobiCN[p, k^2]*JacobiSN[tau, k^2]* JacobiCN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[t
k^2]^2)
Protect[JacobiDN]

Unprotect[EllipticE]
(*E(p+tau)*)
EllipticE[JacobiAmplitude[p + tau, k^2], k^2] := EllipticE[JacobiAmplitude[p, k^2], k^2] +
EllipticE[JacobiAmplitude[tau, k^2], k^2] - k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[p
+ tau, k^2]
(*E(-p+tau)*)
```

```
EllipticE[JacobiAmplitude[-p + tau, k^2], k^2] := -EllipticE[JacobiAmplitude[p, k^2], k^2]
+ EllipticE[JacobiAmplitude[tau, k^2], k^2] + k^2*JacobiSN[p, k^2]*JacobiSN[tau, k^2]*JacobiSN[-
p + tau, k^2]
Protect[EllipticE]

Unprotect[JacobiSN]
(*sn(p+tau)*)
JacobiSN[p + tau, k^2] := (JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] + Ja-
cobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*sn(-p+tau)*)
JacobiSN[-p + tau, k^2] := (-JacobiSN[p, k^2]*JacobiCN[tau, k^2]*JacobiDN[tau, k^2] +
JacobiCN[p, k^2]*JacobiDN[p, k^2]*JacobiSN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
Protect[JacobiSN]

Unprotect[JacobiCN]
(*cn(p+tau)*)
JacobiCN[p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] - JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
(*cn(-p+tau)*)
JacobiCN[-p + tau, k^2] := (JacobiCN[p, k^2]*JacobiCN[tau, k^2] + JacobiSN[p, k^2]*JacobiDN[p,
k^2]*JacobiSN[tau, k^2]* JacobiDN[tau, k^2])/(1 - k^2*JacobiSN[p, k^2]^2*JacobiSN[tau,
k^2]^2)
Protect[JacobiCN]

(*Definitions of (x, y, z) coordinates of Extremal Trajectories Function*)
phit[phhi0_, t_] := t + phhi0;
w[phhi0_, k_] := 1/(dn[phhi0, k] - k cn[phhi0, k]);
x11[phhi0_, k_] := w[phhi0, k] + 1/(w[phhi0, k] (1 - k^2));
x12[phhi0_, k_] := k/(w[phhi0, k] (1 - k^2)) - k w[phhi0, k];
y11[phhi0_, k_] := w[phhi0, k] - 1/(w[phhi0, k] (1 - k^2));
y12[phhi0_, k_] := k w[phhi0, k] + k/(w[phhi0, k] (1 - k^2));
x[phhi0_, k_, t_] := 1/2 (x11[phhi0, k] (EE[phit[phhi0, t], k] - EE[phhi0, k]) + x12[phhi0,
k] (sn[phit[phhi0, t], k] - sn[phhi0, k]));
y[phhi0_, k_, t_] := 1/2 (y11[phhi0, k] (EE[phit[phhi0, t], k] - EE[phhi0, k]) - y12[phhi0,
k] (sn[phit[phhi0, t], k] - sn[phhi0, k]));
z[phhi0_, k_, t_] := Log[w[phhi0, k] (dn[phit[phhi0, t], k] - k cn[phit[phhi0, t], k])];

(* Computing x, y, z in coordinates u1, u2, k *)
```

```
u1rules := {JacobiAmplitude[p, k^2] -> u1, JacobiSN[p, k^2] -> Sin[u1], JacobiCN[p, k^2]
-> Cos[u1], JacobiDN[p, k^2] -> Sqrt[1 - k^2 Sin[u1]^2]};
u2rules := {JacobiAmplitude[tau, k^2] -> u2, JacobiSN[tau, k^2] -> Sin[u2], JacobiCN[tau,
k^2] -> Cos[u2], JacobiDN[tau, k^2] -> Sqrt[1 - k^2 Sin[u2]^2]};


x[phhi0, k, t] /. {phhi0 + t -> tau + p} /. {phhi0 -> tau - p} /. u1rules /. u2rules;
Simplify[%];
(*Take the result of simplification and define a new function \ xu=result of last statement*)

y[phhi0, k, t] /. {phhi0 + t -> tau + p} /. {phhi0 -> tau - p} /. u1rules /. u2rules;
Simplify[%];
(*Take the result of simplification and define a new function \ yu=result of last statement*)

z[phhi0, k, t] /. {phhi0 + t -> tau + p} /. {phhi0 -> tau - p} /. u1rules /. u2rules;
Simplify[%];
(*Take the result of simplification and define a new function \ zu=result of last statement*)

(* Defining x, y, z in coordinates u1, u2, k from the expressions \ computed above *)


zu[u1_, u2_, k_] := Log[(1 + k Sin[u1] Sin[u2])/(1 - k Sin[u1] Sin[u2])]


yu[u1_, u2_, k_] := (k (-(-8 + 2 k^2 - 2 k^2 Cos[2 u1] + k^2 Cos[2 (u1 - u2)] - 2 k^2
Cos[2 u2] + k^2 Cos[2 (u1 + u2)]) EllipticE[u1, k^2] (k^3 - 2 Cos[u1] Sqrt[2 - k^2 + k^2
Cos[2 u1]] Cos[u2] Sqrt[ 2 - k^2 + k^2 Cos[2 u2]] + k Cos[u1]^2 (2 - k^2 + k^2 Cos[2
u2]) - 2 k Sin[u1]^2 + k^3 Sin[u1]^2 - k Cos[u2]^2 (-2 + k^2 + k^2 Sin[u1]^2) - 4 Sin[u1]
Sin[u2] + 4 k^2 Sin[u1] Sin[u2] - 2 k Sin[u2]^2 + k^3 Sin[u2]^2 + k^3 Sin[u1]^2 Sin[u2]^2)
+ 2 Sqrt[2] Sin[u1] (2 + 2 k Sin[u1] Sin[u2]) (-k^2 Cos[u1]^2 Cos[u2] Sqrt[ 2 - k^2 + k^2
Cos[2 u2]] (-2 + 3 k Sin[u1] Sin[u2]) + Cos[u2] Sqrt[ 2 - k^2 + k^2 Cos[2 u2]] (4 - 2 k^2
- 2 k^2 Sin[u1]^2 - k^3 Sin[u1] Sin[u2] + k^3 Sin[u1]^3 Sin[u2]) + k Cos[u1] Sqrt[ 2 - k^2
+ k^2 Cos[2 u1]] (1/ 2 k (8 - 7 k^2 + k^2 Cos[2 u2]) Sin[u1] Sin[u2] + Cos[u2]^2 (-4 + 3
k^3 Sin[u1] Sin[u2])))))/(16 (-1 + k^2) (-1 + k Sin[u1] Sin[u2]) (1 + k Sin[u1] Sin[u2])^2 (k
Cos[u1] Cos[u2] - Sqrt[1 - k^2 Sin[u1]^2] Sqrt[1 - k^2 Sin[u2]^2]))


xu[u1_, u2_, k_] := ((8 - 2 k^2 + 2 k^2 Cos[2 u1] - k^2 Cos[2 (u1 - u2)] + 2 k^2 Cos[2
u2] - k^2 Cos[2 (u1 + u2)]) EllipticE[u1, k^2] (-4 + 3 k^2 - 3 k^2 Cos[u1]^2 Cos[u2]^2 +
2 k Cos[u1] Sqrt[2 - k^2 + k^2 Cos[2 u1]] Cos[u2] Sqrt[ 2 - k^2 + k^2 Cos[2 u2]] + k^2
Cos[u2]^2 Sin[u1]^2 + 4 k (-1 + k^2) Sin[u1] Sin[u2] + k^2 Sin[u2]^2) + 2 Sqrt[2] k^2 Sin[
u1] (2 + 2 k Sin[u1] Sin[u2]) (1/2 Cos[u1] Sqrt[ 2 - k^2 + k^2 Cos[2 u1]] (8 - 4 k^2 - k
```

```
Cos[u1 - 3 u2] - k Cos[u1 - u2] + 4 k^2 Cos[2 u2] + k Cos[u1 + u2] + k Cos[u1 + 3 u2])
+ k Cos[u1]^2 Cos[u2] Sqrt[ 2 - k^2 + k^2 Cos[2 u2]] (-2 + 3 k Sin[u1] Sin[u2]) - Cos[u2]
Sqrt[ 2 - k^2 + k^2 Cos[2 u2]] (2 k - 2 k Sin[u1]^2 + (-4 + 3 k^2) Sin[u1] Sin[u2] + k^2
Sin[u1]^3 Sin[u2]))/(16 (-1 + k^2) (-1 + k Sin[u1] Sin[u2]) (1 + k Sin[u1] Sin[u2])^2 (k
Cos[u1] Cos[u2] - Sqrt[1 - k^2 Sin[u1]^2] Sqrt[1 - k^2 Sin[u2]^2]))

(* Computing R1, R1 in coordinates u1, u2, k *)
R1u[u1_, u2_, k_] := yu[u1, u2, k] Cosh[zu[u1, u2, k]/2] - xu[u1, u2, k] Sinh[zu[u1, u2,
k]/2];

R2u[u1_, u2_, k_] := yu[u1, u2, k] Sinh[zu[u1, u2, k]/2] + xu[u1, u2, k] Cosh[zu[u1, u2,
k]/2];

u1pk[p_, k_] := JacobiAmplitude[p, k^2];
```

Plots for various R

```
R := 1;
(* s1 = 1 *)
plotr1s10 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], R2u[u1pk[R/2, k], u2, k], zu[u1pk[R/2,
k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, Mesh -> 5, PlotRange -> Automatic, AxesLabel -> {R1,
R2, z}];
(* s1 = - 1 *)
plotr1s11 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], - R2u[u1pk[R/2, k], u2, k], -
zu[u1pk[R/2, k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, Mesh -> 5, PlotRange -> Automatic,
AxesLabel -> {R1, R2, z}];

R := 2;
(* s1 = 1 *)
plotr2s10 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], R2u[u1pk[R/2, k], u2, k], zu[u1pk[R/2,
k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, PlotRange -> Automatic, AxesLabel -> {R1, R2, z}];
(* s1 = - 1 *)
plotr2s11 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], - R2u[u1pk[R/2, k], u2, k], -
zu[u1pk[R/2, k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, PlotRange -> Automatic, AxesLabel
-> {R1, R2, z}];

R := 3;
(* s1 = 1 *)
```

```
plotr3s10 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], R2u[u1pk[R/2, k], u2, k], zu[u1pk[R/2, k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, PlotRange -> Automatic, AxesLabel -> {R1, R2, z}];
(* s1 = - 1 *)
plotr3s11 = ParametricPlot3D[{R1u[u1pk[R/2, k], u2, k], - R2u[u1pk[R/2, k], u2, k], - zu[u1pk[R/2, k], u2, k]}, {k, 0, 1}, {u2, 0, 2 Pi}, PlotRange -> Automatic, AxesLabel -> {R1, R2, z}];
```

D.2 Plot of Part of the Wavefront for Time t = R Corresponding to $\lambda \in C_2$

Setup

```
Clear[x, y, z, zu, yu, zu, R1u, R2u]
w[phhi0_, k_] := 1/(dn[phhi0, k] - k cn[phhi0, k]);
x11[phhi0_, k_] := -w[phhi0, k] + 1/(w[phhi0, k] (1 - k^2));
x12[phhi0_, k_] := k/(w[phhi0, k] (1 - k^2)) + k w[phhi0, k];
y11[phhi0_, k_] := w[phhi0, k] + 1/(w[phhi0, k] (1 - k^2));
y12[phhi0_, k_] := k w[phhi0, k] - k/(w[phhi0, k] (1 - k^2));
x[phhi0_, k_, t_] := 1/2 (x11[phhi0/k, k] (EE[phit[phhi0/k, t/k], k] - EE[phhi0/k, k] - (1 - k^2) (phit[phhi0/k, t/k] - phhi0/k)) + x12[phhi0/k, k] (sn[phit[phhi0/k, t/k], k] - sn[phhi0/k, k]));
y[phhi0_, k_, t_] := 1/2 (-y11[phhi0/k, k] (EE[phit[phhi0/k, t/k], k] - EE[phhi0/k, k] - (1 - k^2) (phit[phhi0/k, t/k] - phhi0/k)) + y12[phhi0/k, k] (sn[phit[phhi0/k, t/k], k] - sn[phhi0/k, k]));
z[phhi0_, k_, t_] := Log[w[phhi0/k, k] (dn[phit[phhi0/k, t/k], k] - k cn[phit[phhi0/k, t/k], k])];


x[phhi0, k, t] /. {phhi0/k + t/k -> tau + p} /. {phhi0/k -> tau - p} /. u1rules /. u2rules;
Simplify[%];
y[phhi0, k, t] /. {phhi0/k + t /k -> tau + p} /. {phhi0 /k -> tau - p} /. u1rules /. u2rules;
Simplify[%];
z[phhi0, k, t] /. {phhi0 /k + t/k -> tau + p} /. {phhi0/k -> tau - p} /. u1rules /. u2rules;
Simplify[%];


zu[u1_, u2_, k_, t_] := Log[(1 + k Sin[u1] Sin[u2])/(1 - k Sin[u1] Sin[u2])]


yu[u1_, u2_, k_, t_] := ((-2 + k^2 - k^2 Cos[u1]^2 Cos[u2]^2 + k^2 Cos[u2]^2 Sin[u1]^2 + 2 k (-1 + k^2) Sin[u1] Sin[u2] + k^2 Sin[u2]^2 + 2 k Cos[u1] Cos[u2] Sqrt[1 - k^2 Sin[u1]^2]
```

```
Sqrt[ 1 - k^2 Sin[u2]^2]) (t - k^2 t + k^2 (-1 + k^2) t Sin[u1]^2 Sin[u2]^2 + k^3 Sqrt[1 - k^2 Sin[u1]^2] Sin[2 u1] Sin[u2]^2 + 2 k EllipticE[u1, k^2] (-1 + k^2 Sin[u1]^2 Sin[u2]^2)) - 2 k^3 Cos[u2] Sin[u1] Sqrt[ 1 - k^2 Sin[u2]^2] (-k (1 + Cos[u1]^2) Cos[u2]^2 + 2 Cos[u1] Cos[u2] Sqrt[1 - k^2 Sin[u1]^2] Sqrt[ 1 - k^2 Sin[u2]^2] + Sin[u1] (-2 (-1 + k^2) Sin[u2] + Sin[u1] (k + (k - 2 k^3) Sin[u2]^2))))/(2 k (-1 + k^2) (-1 + k Sin[u1] Sin[u2]) (1 + k Sin[u1] Sin[u2])^2 (k Cos[u1] Cos[u2] - Sqrt[1 - k^2 Sin[u1]^2] Sqrt[1 - k^2 Sin[u2]^2]))

xu[u1_, u2_, k_, t_] := (2 k Cos[u2] Sin[u1] Sqrt[ 1 - k^2 Sin[u2]^2] (-2 + k^2 - k^2 Cos[u1]^2 Cos[u2]^2 + k^2 Cos[u2]^2 Sin[u1]^2 + 2 k (-1 + k^2) Sin[u1] Sin[u2] + k^2 Sin[u2]^2 + 2 k Cos[u1] Cos[u2] Sqrt[1 - k^2 Sin[u1]^2] Sqrt[ 1 - k^2 Sin[u2]^2]) + (t - k^2 t + k^2 (-1 + k^2) t Sin[u1]^2 Sin[u2]^2 + k^3 Sqrt[1 - k^2 Sin[u1]^2] Sin[2 u1] Sin[u2]^2 + 2 k EllipticE[u1, k^2] (-1 + k^2 Sin[u1]^2 Sin[u2]^2)) (k (1 + Cos[u1]^2) Cos[ u2]^2 - 2 Cos[u1] Cos[u2] Sqrt[1 - k^2 Sin[u1]^2] Sqrt[ 1 - k^2 Sin[u2]^2] + Sin[u1] (2 (-1 + k^2) Sin[u2] + k Sin[u1] (-1 + (-1 + 2 k^2) Sin[u2]^2))))/(2 (-1 + k^2) (-1 + k Sin[u1] Sin[u2]) (1 + k Sin[u1] Sin[u2])^2 (k Cos[u1] Cos[u2] - Sqrt[1 - k^2 Sin[u1]^2] Sqrt[1 - k^2 Sin[u2]^2]))

(* Computing R1, R2 in coordinates u1, u2, k *)
R1u[u1_, u2_, k_, t_] := yu[u1, u2, k, t] Cosh[zu[u1, u2, k, t]/2] - xu[u1, u2, k, t] Sinh[zu[u1, u2, k, t]/2];
R2u[u1_, u2_, k_, t_] := yu[u1, u2, k, t] Sinh[zu[u1, u2, k, t]/2] + xu[u1, u2, k, t] Cosh[zu[u1, u2, k, t]/2];
u1pk[p_, k_] := JacobiAmplitude[p, k^2];
```

Plots

```
R := 1;
(* s2 = 1 *)
plotr1s20 = ParametricPlot3D[{R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2, k, R], zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, AxesLabel -> {R1, R2, z}];
(* s2 = - 1 *)
plotr1s21 = ParametricPlot3D[{- R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2, k, R], - zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, AxesLabel -> {R1, R2, z}];
R := 2;
(* s2 = 1 *)
plotr2s20 = ParametricPlot3D[{R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2, k, R], zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, Mesh -> 5, AxesLabel -> {R1, R2, z}];
(* s2 = - 1 *)
```

```
plotr2s21 = ParametricPlot3D[{- R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2,
k, R], - zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, AxesLabel -> {R1, R2, z}];
R := 3;
(* s2 = 1 *)
plotr3s20 = ParametricPlot3D[{R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2,
k, R], zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, AxesLabel -> {R1, R2, z}];
(* s2 = - 1 *)
plotr3s21 = ParametricPlot3D[{- R1u[u1pk[R/(2*k), k], u2, k, R], R2u[u1pk[R/(2*k), k], u2,
k, R], - zu[u1pk[R/(2*k), k], u2, k, R]}, {k, 0, 1}, {u2, 0, 2 Pi}, AxesLabel -> {R1, R2, z}];


(*Plots of planes R1 = 0, R2 = 0 and z = 0 for Cutouts*)
plotz = ParametricPlot3D[{x,y,0},{x,-10,10},{y,-10,10},AxesLabel->{x,y,z}];
plotr1 = ParametricPlot3D[{0,y,x},{x,-10,10},{y,-10,10},Mesh->5,AxesLabel->{x,y,z}];
plotr2 = ParametricPlot3D[{x,0,y},{x,-10,10},{y,-10,10},Mesh->5,AxesLabel->{x,y,z}];
```

D.3 Complete Plot

```
(*Sub-Riemannian Wavefront*)
Show[plotr2s10, plotr2s11, plotr2s20, plotr2s21, PlotRange -> All, AspectRatio -> 1, Plot-
Label -> Text["Wavefront for R=2"]]
(*Cutout of sub-Riemannian Wavefront*)
(*This produces Figure 9.8*)
Show[plotr2s10,plotr2s11,plotr2s20,plotr2s21, plotz, plotr1,PlotRange->{All,{-0.5,5},All}, AspectRatio-
>2,PlotLabel->Text["Wavefront Cutout for R=2"]]
Show[plotr2s10,plotr2s11,plotr2s20,plotr2s21, plotz, plotr1,PlotRange->{All,{-0.5,0.5},All},
AspectRatio->2,PlotLabel->Text["Wavefront Cutout for R=2"]]
(*Matryoshka of sub-Riemannian Wavefront*)
Show[plotr1s10, plotr1s11, plotr1s20, plotr1s21, plotr2s10, \ plotr2s11, plotr2s20, plotr2s21,
plotr3s10, plotr3s11, plotr3s20, \ plotr3s21, PlotRange -> {All, {-0.5, All}, All}, AspectRatio
-> 0.6]
(*Matryoshka of sub-Riemannian Sphere*)
Show[plotr1s10, plotr1s11, plotr1s20, plotr1s21, plotr2s10, \ plotr2s11, plotr2s20, plotr2s21,
plotr3s10, plotr3s11, plotr3s20, \ plotr3s21, PlotRange -> {All, {1, 20}, All}, AspectRatio
-> ]
```

Bibliography

[1] I. Moiseev, Yu. L. Sachkov. Maxwell strata in sub-Riemannian problem on the group of motions of a plane. *ESAIM: COCV*, 16:380–399, 2010.

[2] A. A. Agrachev, D. Barilari. Sub-Riemannian structures on 3D Lie groups. *Journal of Dynamical and Control Systems*, 18(1):21–44, 2012.

[3] Y. A. Butt, Yu. L. Sachkov, A. I. Bhatti. Extremal trajectories and Maxwell strata in sub-Riemannian problem on group of motions of pseudo-Euclidean plane. *Journal of Dynamical and Control Systems*, 20(3):341–364, July 2014.

[4] Y. A. Butt, A. I. Bhatti, Yu. L. Sachkov. Integrability by quadratures in optimal control of a unicycle on hyperbolic plane. Chicago Illionis, 1–3, Jul 2015. American Control Conference.

[5] Y. A. Butt, Yu L. Sachkov, A. I. Bhatti. Maxwell strata and conjugate points in the sub-Riemannian problem on the Lie group SH(2). *arXiv:1408.2043v1*, 2014.

[6] R. Sjamaar. Manifolds and differential forms. Lecture Notes Math 321, Ithaca, New York,, 2006.

[7] V. G. Ivancevic, T. T. Ivancevic. *Lecture Notes in Lie Groups.* arXiv:1104.1106v2, 2011.

[8] D. D. Nolte. The tangled tale of phase space. *Physics Today - American Institute of Physics*, April 2010.

[9] J. M. Lee. *Introduction to Topological Manifolds.* Springer, 2011.

[10] J. M. Lee. *Introduction to Smooth Manifolds.* Springer, 2011.

[11] V. Jurdjevic. *Geometric Control Theory.* Cambridge University Press, 1997.

[12] V. G. Ivancevic, T. T. Ivancevic. *Applied Differnetial Geometry, A Modern Introduction.* World Scientific Publishing Co. Pte. Ltd., 2007.

[13] V. I. Arnold. On teaching mathematics. *Russian Math. Surveys*, 1998.

[14] E. F. Robertson T. S. Blyth. *Groups, rings and fields: Algebra through practice.* Cambridge University Press, 1985.

[15] Yu.L. Sachkov. Control theory on Lie groups. *Journal of Mathematical Sciences*, 156(3):381–439, 2009.

[16] S. M. LaValle. *Planning Algorithms.* Cambridge University Press, 2006.

[17] A.M. Bloch, J. Baillieul, P. Crouch, J. Marsden. *Nonholonomic Mechanics and Control, Interdisciplinary Applied Mathematics*, volume 24. Springer, 2007.

[18] M. Gromov. *Carnot-Caratheodory spaces seen from within.* Math, 1996.

[19] J. M. Lee. *Riemannian Manifolds - An Introduction to Curvature.* Springer, 1997.

[20] A. A. Agrachev, Yu. L. Sachkov. *Control Theory from the Geometric Viewpoint.* Springer, 2004.

[21] R. Penrose. *The Road to Reality - A Complete Guide to the Laws of the Universe.* Jonathan Cape London, 2004.

[22] Euclid Translated by Sir Thomas Heath. *The Elements: Books I-XIII-Complete and Unabridged.* Barnes & Noble ISBN 0-7607-6312-7, 2006.

[23] Sir Isaac Newton. *Philosophiae Naturalis Principia Mathematica, English Translation.* Cambridge, Mass. Harvard University Press, 1833.

[24] M. Hazewinkel. *Lobachevskii Geometry, Encyclopedia of Mathematics.* Springer, 2001.

[25] M. Hazewinkel. *Elliptic Geometry, Encyclopedia of Mathematics.* Springer, ISBN 978-1-55608-010-4, 1993.

[26] M. D. Tellez, A. Ibort. A panorama of geometrical optimal control theory. *Extracta Mathematica*, 18:129–151, 2003.

[27] I. M. Ross. *The Brachistochrone Paridgm, in A Primer on Pontryagin's Principle in Optimal Control.* Collegiate Publishers, 2009.

[28] H. J. Sussman, J. C. Willems. 300 years of optimal control: from the brachystochrone to the maximum principle. *IEEE Control Systems*, 17, 1997.

[29] R. V. Gamkrelidze. Discovery of the maximum principle in optimal control. *Journal of Dynamical and Control Systems*, 5, 1999.

[30] E. Ortacgil. The heritage of S. Lie and F. Klein: Geometry via transformation groups. *arXiv:math/0604223v3*, 2013.

[31] X. Hu, A. Lindquist. *Geometric Control Theory.* Lecture Notes of the Royal Institute of Technology, Stockholm, 2011.

[32] E. Todorov. *Optimal Control Theory - In Bayesian Brain: Probabilistic Approaches to Neural Coding, chap 12.* MIT Press, 2006.

[33] A. D. Lewis. *The Maximum Principle of Pontryagin in control and in optimal control.* 2006.

[34] G. Welch, G. Bishop. *An Introduction to the Kalman Filter - Lecture Notes.* Department of Computer Science, University of North Carolina at Chapel Hill, 2006.

[35] R. Bellman. *Dynamic Programming.* Princeton University Press, 1957.

[36] R. P. Kumar. Dynamic programming. *Control Systems Robotics and Automation*, 2001.

[37] R. Montgomery. *A tour of sub-Riemannian geometries, their geodesics and applications.* Number 91 in Mathematical Surveys and Monographs. American Mathematical Society, 2002.

[38] A. A. Agrachev, Davide Barilari, Ugo Boscain. *Introduction to Riemannian and Sub-Riemannian geometry (from Hamiltonian viewpoint).* Preprint SISSA, September 13, 2012.

[39] P. Petersen. *Riemannian Geometry*, volume 171. Springer, second edition, 2006.

[40] E. Le. Donne. Lecture notes on sub-Riemannian geometry. *Preprint*, 2010.

[41] A. M. Vershik, V. Ya. Gershkovich. *Dynamical Systems VII, Integrable Systems, Nonholonomic Dynamical Systems - Chap 1 - Nonholonomic Dynamical Systems,Geometry of Distributions and Variational Problems.* Springer-Verlag, 2007.

[42] J. Petitot. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *Journal of Physiology - Paris*, 97:265–309, 2003.

[43] A. P. Mashtakov, A. A. Ardentov, Yu. L. Sachkov. Parallel algorithm and software for image inpainting via sub-Riemannian minimizers on the group of rototranslations. *Numerical Mathematics Theory Methods and Applications*, 6:95–115, 2013.

[44] R. Montgomery. Isoholonomic problems and some applications. *Communication in Mathematical Physics*, 128:565–592, 1990.

[45] F. Monroy, A. Anzaldo-Meneses. Optimal control on the Heisenberg group. *Journal of Dynamical and Control System*, 5(4):473–499, 1999.

[46] Yu. L. Sachkov. The Maxwell set in the generalized Dido problem. *Sbornik: Mathematics*, 197(4), 2006.

[47] Yu L. Sachkov. Discrete symmetries in the generalized Dido problem. *Sbornik: Mathematics*, 197(2), 2006.

[48] Yu.L. Sachkov. Conjugate and cut time in the sub-Riemannian problem on the group of motions of a plane. *ESAIM: COCV*, 16:1018–1039, 2010.

[49] Yu. L. Sachkov. Cut locus and optimal synthesis in the sub-Riemannian problem on the group of motions of a plane. *ESAIM: COCV*, 17:293–321, 2011.

[50] U. Boscain, F. Rossi. Invariant Carnot-Caratheodory metrics on s^3, $so(3)$, $sl(2)$ and lens spaces. *SIAM, Journal on Control and Optimization*, 47:1851–1878, 2008.

[51] Yu. L. Sachkov. Controllability of right-invariant systems on solvable Lie groups. *Journal of Dynamical and Control Systems*, 3(4), 1997.

[52] A. D. Mazhitova. Sub-Riemannian geodesics on the three-dimensional solvable non-nilpotent Lie group Solv^-. *Journal of Dynamical and Control Systems*, pages 1–14, 2012.

[53] A. A. Ardentov, Yu. L. Sachkov. Extremal trajectories in a nilpotent sub-Riemannian problem on the Engel group. *Sbornik: Mathematics*, 202(11), 2011.

[54] T. Lee, M. Leok, N. H. McClamroch. Geometric tracking control of a quadrotor UAV for extreme maneuverability. In *18th IFAC World CongressMilano (Italy)*, 2011.

[55] T. Lee. *Computational Geometric Mechanics and Control of Rigid Bodies*. PhD thesis, University of Michigan, 2008.

[56] R. LaBudde, D. Greenspan. Energy and momentum conserving methods of arbitrary order for the numerical integration of equations of motion - motion of a single particle. *Numerische Mathematik*, 25(4):323–346, 1976.

[57] David Thomas. *Modern Geometry*. Brooks/Cole: Pacific Grove, CA., 2002.

[58] Y. C. Wong. Euclidean n-planes in pseudo-Euclidean spaces and differential geometry of Cartan domains. *Bulletin of the American Mathematical Society*, 75(2):409–414, 1969.

[59] F. Catoni, D. Boccaletti, R. Cannata, V. Catoni, E. Nichelatti, P. Zampetti. *The Mathematics of Minkowskian Space-Time with and Introduction to Commutative Hypercomplex Numbers*. Birkhauser Verlag AG, 2008.

[60] N. Ja. Vilenkin. *Special Functions and Theory of Group Representations (Translations of Mathematical Monographs)*. American Mathematical Society, revised edition, 1968.

[61] J. Milnor. *Hyperbolic geometry: the first 150 years*, volume 6. Bulletin of the American Mathematical Society - Lecture Notes of the Royal Institute of Technology, Stockholm, 1982.

[62] A. A. Ungar. Einstein's special relativity : The hyperbolic geometric viewpoint. In *Conference on Mathematics, Physics and Philosophy on the Interpretations of Relativity, II Budapest*, September 2009.

[63] P. Panyakeow and M. Mesbahi. Decentralized deconfliction algorithms for unicycle UAVs. In *American Control Conference (ACC)*, pages 794–799, June 2010.

[64] A. A. Ardentov, Yu. L. Sachkov. Cut time in sub-Riemannian problem on Engel group. *ESAIM:COCV*, 2015.

[65] S. Jacquet. Regularity of sub-Riemannian distance and cut locus. *Univ. Stud. Firenze, Florence, Italy*, (Preprint 35), May 1999.

[66] P. K. Rashevsky. About connecting two points of complete nonholonomic space by admissible curve. *Uch Zapiski Ped*, pages 83–94, 1938.

[67] W. L. Chow. Uber systeme von linearen partiellen dierentialgleichungen erster ordnung. *Math*, pages 98–105, 1939.

[68] W. Magnus. On the exponential solution of differential equations for a linear operator. *Communications on Pure and Applied Math*, 7:649–673, 1954.

[69] J. Wei, E. Norman. On global representations of the solution of linear differential equations as a product of exponentials. *Proceedings of the American Mathematical Society*, 15:327–334, 1964.

[70] H. K. Struemper. *Motion Control for Nonholonomic Systems on Matrix Lie Groups*. PhD thesis, Department of Electrical Engineering, University of Maryland, 1997.

[71] E. T. Whittaker, G. N. Watson. *A Course of Modern Analysis, An introduction to the general theory of infinite processes and of analytic functions; with an account of principal transcendental functions*. Cambridge University Press, 1996.

[72] Yu. L. Sachkov. Closed Euler elasticae. *Proceedings of the Steklov Institute of Mathematics*, 28, 2012.

[73] I. S. Gradshteyn, I. M. Ryzhik. *Table of Integrals, Series, and Products*, volume 2007. Academic Press, 7 edition, 2007.

[74] R. S. Palais. *A Modern Course on Curves and Surfaces*. Virtual Math Museum, 2003.

[75] Stanislaw H. Zak. *Systems and Control*. Oxford University Press, 2003.

[76] A. A. Agrachev. *Introduction to optimal control theory.* in Mathematical control theory, Part 1, 2, ICTP Lecture Notes, VIII, Abdul Salam International Centre of Theoretical Physics, (electronic), 2002.

[77] Yu. L. Sachkov. Complete description of the Maxwell strata in the generalized Dido problem. *English translation in Sbornik Mathematics*, pages 901–950, 2006.

[78] Yu. L. Sachkov. Maxwell strata in the Euler elastic problem. *Journal of Dynamical and Control Systems*, 14(2):169–234, April 2008.

[79] V. I. Arnold. *Singularities of Caustics and Wave Fronts (Mathematics and its Applications)*, volume 62. Springer, 1991.

[80] O. Bolza. *Lecture on Calculus of Variations.* The University of Chicago Press, 1904.

[81] Yu. L. Sachkov. Conjugate points in the Euler elastic problem. *Journal of Dynamical and Control Systems*, 14:409–439, 2008.

[82] A. A. Ardentov, Yu. L. Sachkov. Conjugate points in nilpotent sub-Riemannian problem on the Engel group. *Journal of Mathematical Sciences*, 195:369–390, 2013.

[83] A. A. Agrachev. Geometry of optimal control problems and Hamiltonian systems. *Nonlinear and Optimal Control Theory, Lecture Notes in Mathematics. CIME, 1932, Springer Verlag*, pages 1–59, 2008.

[84] J. V. Armitage, W. F. Eberlein. *Elliptic Functions.* Cambridge University Press, 2006.

[85] S. Wolfram. *Mathematica: a system for doing mathematics by computer.* Addison-Wesley,, 1992.

[86] A. S. Poznyak. *Advanced Mathematical Tools for Automatic Control Engineers Volume 1: Deterministic Techniques.* Elsevier, 2008.