Finsler Optimal Control Theory

by

Srikanth Gottipati

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York.

2009

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date ______ Sergei Artemov

Chair of Examining Committee

Date ______ Theodore Brown

Executive Officer

Anil Nerode

Robert Haralick

Rohit Parikh

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Finsler Optimal Control Theory

by

Srikanth Gottipati

Adviser: Professor Sergei Artemov

This thesis is a contribution to solving problems of extracting optimal controls for complex systems. Its novelty consists of a detailed examination of Finsler geometry based control by connections as proposed by Kohn and Nerode and its relation to Pontryagin's maximum principle. The long term hope is that these methods will form the underpinning of the applications of control of hybrid systems.

Any advances in mathematical and algorithmic techniques for solving such problems would have wide application in business, industry, and science. The widespread use of Bellman's dynamic programming, Dantzig's linear programming, Kalman's optimization with linear quadratic cost functions, demonstrate this. But symbolic and numerical techniques historically have fallen well short of yielding efficient computation procedures to obtain near optimal controls for complex systems. Most investigations have been based on Pontryagin's geometric representation of optimal control and his maximum principle. Kohn and Nerode have proposed a different problem formulation aimed at extracting robust controls as a function of state (synthesis problem with a robustness requirement) in the context of a Finsler geometry corresponding to the optimal control problem. This leads to the study of geometric, symbolic, and numerical methods for solving geodesic equations for connections in Finsler Geometry. A principal result of this thesis is the determination of the relations between the Finsler and Pontryagin formulations of optimal control, and the transformation from one to the other. A second principal result is establishing the relations between robustness and curvature. Curvature is used to quantify the spread of geodesics due to disturbances. Finally, the thesis concludes with numerical integration schemes for computing controls and local connections.

Acknowledgements

I would like to sincerely thank my advisor, Prof. Sergei Artemov, for believing in me, encouraging me to think and act independently, and guiding me through the dissertation process.

I am greatly indebted to Prof. Anil Nerode. This dissertation work would not have been accomplished without his mentorship and support. He has been a constant source of guidance throughout the dissertation process.

I am grateful to Prof. Robert Haralick and Prof. Rohit Parikh for their patience and prompt advice as my dissertation committee members. I would like to thank Prof. Irina Gladkova and Prof. Michael Grossberg for their continual support and encouragement all these years.

I owe my gratitude to my parents for their unconditional support and love all these years. Finally, I would like to thank my wife for her constant support and encouragement during my life as a doctoral student.

Contents

List of Figures

Chapter 1

Introduction

1.1 Hybrid Systems

Hybrid systems are complex dynamic systems composed of an interacting network of subsystems with continuous dynamics (we refer to them as plants) and subsystems with discrete dynamics (we refer to them as digital controllers). The continuous and discrete dynamics interact. Changes occur in the plant states in response to changes in the environment and changes in states of the digital controllers. Changes in the states of the digital controllers occur due to changes in the plant state and the environment. The digital controllers are finite state real time input-output devices, the plant evolutions are governed by differential or difference equations. The hybrid system is the coupled system of discrete and continuous elements (see Figure 1.1).

Hybrid systems arise in many contexts, both in man-made systems and in nature. Continuous systems which have a phased operation, such as walking robots, insect motion, or biological cell growth and division, are well-suited to be modelled as hybrid systems, as are continuous systems which are controlled by a discrete logic, such as a chemical plant controlled with valves and pumps, or the autopilot modes for controlling an aircraft. Hybrid systems are also natural models for systems comprised of many interacting subsystems or processes, such as air or ground transportation systems. In these examples, the continuous dynamics model system state evolution, biochemical or chemical reactions, while the discrete dynamics control the sequence of contacts or collisions in the gait cycle, cell divisions, valves and pumps switching, and coordination protocols. In all of these examples, the system dynamics are complex enough that traditional analysis and control methods based solely on differential equations have not been computationally feasible for complex systems. Another important way in which hybrid systems arise is from the hierarchial organization of complex systems. In these systems, a hierarchial organization helps manage complexity and higher levels in the hierarchy require less detailed models of the functioning of the lower levels, necessitating the interaction of discrete and continuous components. Example of such systems include flexible manufacturing and chemical process control systems, intelligent highway systems, air traffic management systems, etc.

To understand the behavior of hybrid systems, to simulate, and to control these systems, theoretical advances and analytical tools are needed. Recent developments include building theory and analytical tools, and applying them to interesting and large scale systems which are currently not known how to analyze, control, or simulate like an automated air traffic system, biological control circuit, etc.

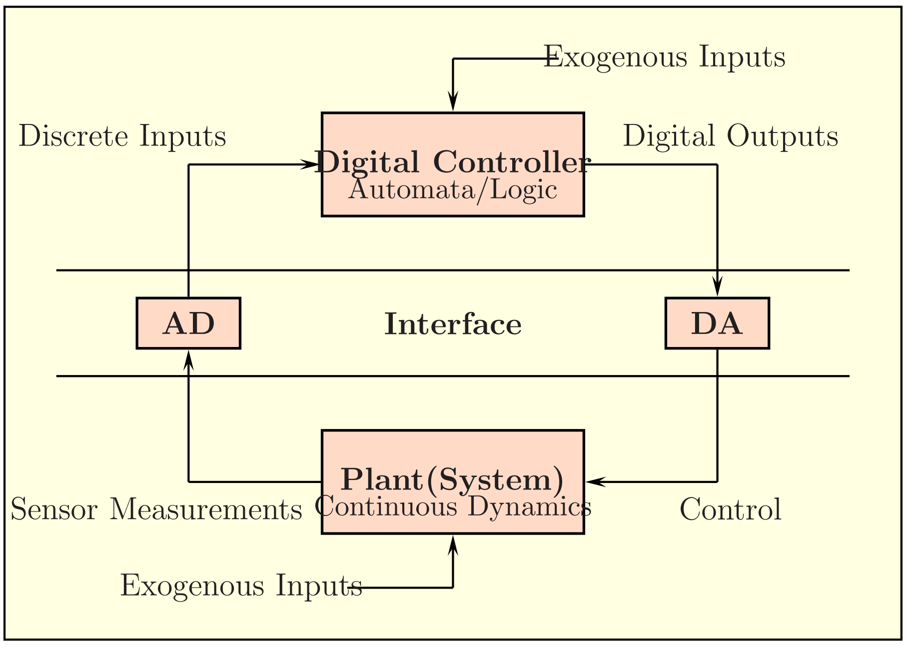

Figure 1.1: Hybrid system.

The predominant model being used in the literature for hybrid systems is to partition the entire state space into finite regions, possibly overlapping, where each region has its own invariant. Once the evolved state of the system in a particular region violates this invariant condition, the state of the system is thrown by the digital control program into a different region with possible reset of the state. From then on the system evolves according to the state dynamics specified for that other region, as long as the invariant for that region is maintained.

The other modeling perspective is that the state space has a finite number of open regions which are represented by logical propositions which involve systems of differential equalities and inequalities. When the state enters one of these regions, one of the logical rules might fire in which case the system dynamics and constraints might change. This in effect is reflected in the change in control policy. When several of the rules fire, there is supposed to be consistent behavior.

The analysis of the behavior of classes of hybrid systems has become quite popular. We are primarily interested in extraction of controls which guarantee epsilon optimal control relative to a cost function and constraints. We are less concerned with taking a given hybrid system (with controls given too) and trying to figure out its behavior as a dynamical system, which is a principal focus of much current research.

1.1.1 History of Control

Here is some background. We first recapitulate some of the evolution of control theory for continuous plants. The modern control of industrial devices by feedback goes back at least to the steam engine governor of James Watt (1787), the feedback linear amplifiers of E. H.Armstrong's patent (1914), the theory of stability of linear feedback amplifier developed at Bell Labs by H. S. Black(1927), H. Nyquist (1932), H. W. Bode (1940) (Laplace transform frequency domain analysis), the theory of linear servomechanisms, developed at MIT Radiation labs, during World War II, emerging in the book "Theory of Servomechnisms" by James, Nichols, and Phillips, 1947. Complex interacting linear systems can be represented in matrix form, giving rise to linear control theory. In 1960, Kalman gave a straightforward matrix calculus algorithm for finding optimal linear controllers for linear models with a linear quadratic cost function when they exist, based on simple facts about matrices and quadratic forms. This remains the principal optimal control tool for the control engineer to this day, especially for process control.

For non-linear optimization in the 1950's Richard Bellman invented dynamic programming for finding paths satisfying constraints and minimizing a cost function assuming his "principle of optimality", based on backwards search for an optimal trajectory starting at the goal. In the 1950's, L.S. Pontryagin developed a differential equation necessary condition for optimal trajectories subject to differential equality and inequality constraints. His maximum principle can be expressed in terms of an associated symplectic geometry. His principle, earlier discovered by Hestenes in 1950, in the context of the calculus of variations, is a local necessary condition for a minimum, just as in calculus setting a derivative equal to zero is a local necessary condition for a minimum. The purpose of local necessary conditions is to cut down the set of controls that have to be considered to find an optimal control. Pontryagin's maximum principle plays for optimality the same role that the first and second derivative tests for minima do in calculus. It is a variant and generalization of the first and second variation necessary tests of the calculus of variations. Most calculus of variation books explore local necessary conditions and local sufficient conditions for minima, but go no further.

What are the mathematical theorems that prove the existence of optimal curves from start to goal? In calculus, we use Weierstrass's sufficiency theorem where the variational operator operates on f and converges to a point where the minimum is achieved (Specker, 1959). But if we say that in practice, we only need to be within epsilon of the minimum, there are algorithms, even efficient algorithms when some smoothness is assumed. The situation is similar in the calculus of variations and optimal control. Hilbert's "direct method" (1901) proves the existence of curves minimizing from start to goal by application of Arzela-Ascoli lemma on uniformly bounded equicontinuous families. Application of this to suitable function spaces produces a compact function space such that, on applying the theorem that an upper semicontinuous function on a compact set attains its minimum, one gets the minimizing curve. Compactness is achieved by extending the space of piecewise smooth curves to a completion using the method first introduced in the calculus of variations by E. J. MacShane (1939-41). These are the "weak solutions", also called relaxed solutions, or measure valued solutions according to the Riesz representation theorem. This was developed further by L.C. Young, Pontryagin, and their followers. Here too algorithms do not exist which will in general compute optimal measure-valued controls. But, just as in calculus, if one asks only for controls which yield trajectories with cost within a prescribed epsilon of optimal cost, there are such algorithms yielding piecewise smooth epsilon optimal control functions, and becomes a question of developing efficient algorithms. Wolf Kohn and Anil Nerode in 1992-4 began the development and implementation of efficient real time algorithms for epsilon optimal control of complex systems. They founded a company, Clearsight Systems, which has demonstrated the real time feasibility of using such chattering approximations to optimal control for many applications such as logistics enterprise systems, financial analysis, and many other areas.

What does this have to do with hybrid systems? The algorithms for real time epsilon optimal control for a fixed epsilon extract a finite state digital controller automaton from the system equations, constraints, and cost function. This automaton senses system state, and on the basis of its state, changes state and issues a chattering control to the actuators to control the system. This was the motivation for Kohn-Nerode defining hybrid automata in the 1992 volume "Hybrid Systems", LNCS, [53]. So hybrid automata and hybrid systems emerge when one extracts epsilon optimal control for continuous non-linear systems and implements it.

There is a second place where the hybrid system concept is important. In many complex systems, behavior is constrained by logical rules. If an order is given by a commander to troops or by a high business administrator to his organization, the course of action must satisfy doctrine, regulations, and other constraints which are expressed in logic. Traditional control theory simply does not deal with this. Those who study planning do deal with this, but primarily by logical deduction of a plan meeting logical constraints and achieving a goal. Prolog is suited to this. But when the system being controlled is a physical system like an airplane or a chemical process plant, we now have purely digital rules in logical form interacting with an evolving physical dynamical system, a hybrid system. The Kohn-Nerode approach converts the logical constraints and goals into (generally non-convex) continuous constraints and goals, and couples the continuous constraints on the controlled physical system with the continualization of the discrete constraints.

Our intention, following Kohn-Nerode [44], [53], is to cut down the search space and to improve the algorithms for computing optimal trajectories using Finsler geometries associated with optimal control problems, specifically to introduce use of Finsler curvature formulas to guide computations. This entails an extensive analysis of Finsler geometry in control theory.

1.1.2 Plant Dynamics

The dynamics of the plant are governed by control systems which are either ordinary differential equations(ODE) or differential-algebraic equations(DAE). A conventional continuous controlled dynamical system is described by the vector field

$$\dot{x} = f(x, u, t), \tag{1.1}$$

which depends on the continuous state x, the continuous control input u, and time t. The above equation represents a non-autonomous system, where, the vector field depends on time t explicitly. However, autonomous control systems are represented as

$$\dot{x} = f(x(t), u(t)), \tag{1.2}$$

where time enters implicitly in the state and control inputs. These control systems given a control function $u(t)$ for a given time interval respect the usual existence and uniqueness theorems of ODEs. However, in some practical applications the control system is given by the following DAE

$$F(x, \dot{x}, u, t) = 0. \tag{1.3}$$

These kind of systems are very difficult to study as the existence and uniqueness theorems of ODEs no longer hold.

In large scale systems however, obtaining the differential equation for the plant is often a difficult task unlike in physical systems where the equations of motion could be easily derived. Hence modeling happens to be one of the crucial issues in Hybrid systems literature. These models have to be updated quite frequently owing to changes in its environment.

1.1.3 Continualization

How does one model the interaction between physical laws governed by differential equations and logical rules which are discrete? Our method is to continualize, that is to replace the discrete constraints and rules by continuous ones. There have been several continualization techniques proposed in the literature of Jeroslow [38] and Williams [74]. One can represent the standard connectives of propositional logic by the following equality or inequality constraints:

$$X_1 \equiv x_1 = 1$$

$$\neg X_1 \equiv x_1 = 0$$

$$X_1 \wedge X_2 \equiv x_1 + x_2 \geq 2$$

$$X_1 \vee X_2 \equiv x_1 + x_2 \geq 1$$

$$X_1 \rightarrow X_2 \equiv x_1 - x_2 \leq 0$$

$$X_1 \leftrightarrow X_2 \equiv x_1 - x_2 = 0$$

along with the additional constraints

$$x_i(1 - x_i) = 0, \;\; x_i \in [0, 1], \;\; i \in \{1, 2\}.$$

There are numerous ways of continualizing other than the above mentioned one, see [27] for more continualization methods. The numerical technique involved in solving the inference problem depends on the particular continualization method chosen.

Williams proposes the following method to transform propositional logic statements into mixed-integer linear inequalities involving integer and continuous variables. This transformation has also been used in the piecewise affine models for hybrid systems, see [11]. In [11], mixed logical dynamical systems were developed by a combination of linear dynamic equations for continuous variables and propositional logic statements and automata for discrete variables. The key idea of this approach consists of embedding the logic part in the state equations by transforming Boolean variables into 0-1 integers, and by expressing the relations as mixed-integer linear inequalities. The following correspondence between a Boolean variable X and its associated binary variable δ will be used: $X = \text{true} \Leftrightarrow \delta = 1$ and $X = \text{false} \Leftrightarrow \delta = 0$. Boolean variables can be defined from linear-threshold conditions over the continuous variables as follows:

$$[X = \text{true}] \leftrightarrow [ax \leq b], \quad x, a \in \mathbb{R}^n, \quad b \in \mathbb{R},$$

$$\text{with} \quad (a'x - b) \in [m, M].$$

This condition can be equivalently represented by the following mixed-integer inequalities:

$$ax - b \leq M(1 - \delta)$$
$$ax - b > m\delta$$

$$\text{IF} \quad X \quad \text{THEN } z = a_1 x - b_1 \quad \text{ELSE} \quad z = a_2 x - b_2,$$

can be expressed as

$$
\begin{aligned}
(m_2 - M_1)\delta + z &\le a_2 x - b_2 \\
(m_1 - M_2)\delta - z &\le -a_2 x + b_2 \\
(m_1 - M_2)(1-\delta) + z &\le a_1 x - b_1 \\
(m_2 - M_1)(1-\delta) - z &\le -a_1 x + b_1
\end{aligned}
$$

where, assuming that $x \in \mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{X}$ is a given bounded set,

$$
M_i \ge \sup_{x\in\mathcal{X}}(a_i x - b_i), \qquad m_i \le \inf_{x\in\mathcal{X}}(a_i x - b_i), \quad i = 1, 2,
$$

are upper and lower bounds, respectively, on $(ax - b)$. Note that the above set of inequalities represents the hybrid product $z = \delta(a_1 x - b_1) + (1-\delta)(a_2 x - b_2)$ between binary and continuous variables.

Example 1. Consider the following system

$$
x(t+1) = \begin{cases} 0.8x(t) + u(t) & \text{if} \quad x(t) \ge 0 \\ -0.8x(t) + u(t) & \text{if} \quad x(t) < 0 \end{cases} \tag{1.4}
$$

where

$$
x(t) \in [-10, 10], \text{ and } \quad u(t) \in [-1, 1].
$$

The condition $x(t) \ge 0$ can be associated with a binary variable $\delta(t)$ such that

$$
[\delta(t) = 1] \leftrightarrow [x(t) \ge 0]
$$

By using the above mentioned transformation we could rewrite it as

$$x(t+1) = 1.6z(t) - 0.8x(t) + u(t)$$

subject to

$$z(t) \leq M\delta(t)$$

$$z(t) \geq m\delta(t)$$

$$z(t) \leq x(t) - m(1 - \delta(t))$$

$$z(t) \geq x(t) - M(1 - \delta(t))$$

1.1.4 Hybrid Automata

There have been various models which have been proposed for hybrid systems. One of the most prevalent model in Hybrid systems literature is that of the Hybrid automaton, see [18], [34], [53]. Another model is that of the switched systems which consist of a family of subsystems and a possibility to switch between them. Such systems were considered in [48]. Another important hybrid systems models are the piecewise affine models, see [11], [12], [17]. These are one of the simplest extensions of linear systems that can model a variety of nonlinear systems arbitrarily well. Discrete-time PWA systems in the presence of linear constraints also allow to completely synthesize feedback controls thus partitioning the state-control space into polyhedral regions. Other models include the timed-automata model [1], and the time-event driven models in manufacturing systems, see [21],[22].

In [34], a hybrid automaton H is defined as a collection

$$H = (Q, X, \Omega_u, \Omega_\mu, f, Inv, G, R).$$

The pair $(x, q) \in X \times Q$, consisting of the continuous state x and the discrete state q, is called the state of H. The possible evolutions of the automaton can be described as follows: Starting at an initial state, $(x_0, q_0) \in X \times Q$, given the inputs $u \in \Omega_u$ and $\mu = \Omega_\mu$, the continuous state x evolves according to $\dot{x} = f(x, q, u)$ while the discrete state remains constant. Continuous evolution can go as long as $(x, q, u, \mu) \in Inv$. If the situation arises during the evolution that $(x, q, u, \mu) \in G$, a transition is possible to a new state. The set of possible new states is given by $R(x, q, u, \mu)$. The transition from the old state (x, q) to the new state (x', q') is called a jump if $x' \neq x$ and a switch if $q' \neq q$. After the transition, there are either new transitions or the continuous evolution resumes and whole process is repeated.

The following example gives a flavor of the continualization technique and the associated hybrid automaton.

Example 2. Let us begin with a simple example of a thermostat controlling the temperature of a room by turning on and off a heater. The real-valued variable x denotes the temperature and the system has two control modes: off and on. When the heater is off the temperature of the room falls according to the differential equation

$$\dot{x} \; = \; -Kx$$

and when on, the temperature rises according to

$$\dot{x} \;=\; K(h - x),$$

where h is a constant.

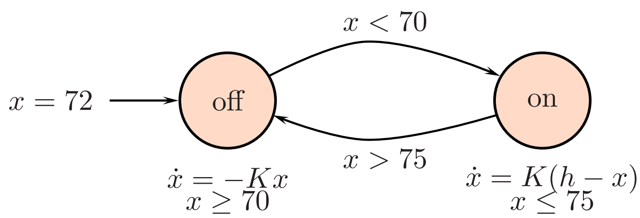

Figure 1.2: Hybrid Automaton: Thermostat

1.2 Optimal Control : Synthesis

Optimal control theory could be viewed as a generalization of Calculus of Variations problems. The problem of optimal plant control is that of constructing a control law or policy $u(t) \in U$, where U is the set of admissible controls, which will drive the plant from its initial state, $x_0 \in M$, into some set of goal states, $G \subset M$, along an optimal path $x(t)$. Optimality is generally formulated as minimizing some cost functional on paths, which we label $J(x)$. In this proposal, the cost functionals will always be assumed to be of the form

$$J(x) := \int_{t_0}^{t_1} L(x(t), u(t))dt, \tag{1.5}$$

where L is the cost function. In practice, moreover, the plant dynamics are constrained to take place on a constraint sub-manifold $N \subset M$. In our work, N is determined by differential-algebraic constraints of the form $F(x, \dot{x}, u) = 0$. These constraints arise from dynamic concerns and from performance specifications. For a particular problem we will have to specify the set of admissible paths from which we will draw candidates for optimality. We shall work with piecewise smooth paths. We shall also assume controllability, i.e., we will assume the existence of a path on the constraint sub-manifold which connects x_0 to G.

In conventional optimal control theory two methods for computing optimal controls stand out. One is the Bellman's *Hamilton-Jacobi-Bellman*(HJB) method and the other is the Pontryagin's *Maximum Principle* method. From an engineer's perspective, the HJB framework is preferred for solving optimal control problems since it naturally leads to a design of optimal *feedback* (or *closed-loop*) controllers, i.e. control laws as a function of state. However, except for some very simple problems, this approach is beset with major theoretical and numerical difficulties. The alternative framework avoids the pitfalls associated with the HJB theory but generates *open-loop* controls, i.e. control laws as a function of time. However, it has been known since the birth of optimal control that if open-loop controls can be generated in real-time, they are basically equivalent to feedback controls. In modern terminology this is *model-predictive control*(MPC) with the horizon being time-to-go. In a later chapter we shall give a full account of the state-of-art methods for computing optimal controls for practical nonlinear systems.

With respect to types of Hybrid models that we described above, they require only open-loop controls if we have full knowledge of the system at the time of computation of optimal controls. With any unforeseen contingencies, for example disturbances or changes in the logic database, one needs to develop feedback controls. Control laws are implemented via a finite state machine. The following simple example shows how the feedback law could be represented by a hybrid automaton.

Double Integrator problem

Consider the control system

$$\frac{d^2x}{dt^2} = u, \tag{1.6}$$

where u is a real control parameter constrained by the condition that $|u| \leq 1$. In the phase coordinates $x^1 = x$ and $x^2 = dx/dt$, this equation may be rewritten in the form of the following system:

$$\frac{dx^1}{dt} = x^2, \qquad \frac{dx^2}{dt} = u. \tag{1.7}$$

Let us consider the problem of getting to the origin $(0, 0)$ from a given initial state x_0 in the shortest time. In other words, we shall consider the time-optimal problem for the case where the origin $(0, 0)$ is the terminal position x_1. The feedback control

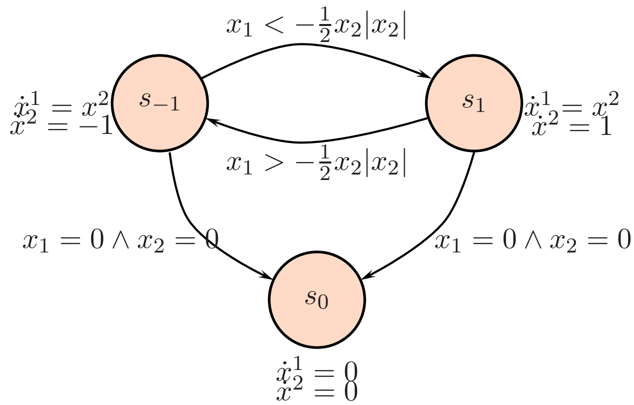

Figure 1.3: Hybrid Automaton: Double Integrator

for the double integrator problem is given by

$$u(t, x_1, x_2) = \begin{cases} -1 & \text{if } x_1 > -\frac{1}{2}x_2|x_2| \\ 1 & \text{if } x_1 > 0 \text{ and } x_1 = -\frac{1}{2}x_2|x_2| \\ 1 & \text{if } x_1 < -\frac{1}{2}x_2|x_2| \\ -1 & \text{if } x_1 < 0 \text{ and } x_1 = -\frac{1}{2}x_2|x_2| \end{cases} \quad (1.8)$$

As seen in the above example the optimal feedback control law partitions the state space to assign optimal controls for each of these partitions. In order to practically implement digital controllers, sensors, actuators, and control logics one needs to quantize the state and/or control space in such a way that the derived control laws upon quantization closely approximate the continuous optimal control laws derived from solving the optimal control problem. Another purpose of the quantization could be regarded as a reduction in the complexity of the control task, mainly in terms of computation and communication requirements.

The quantization problem could formally be posed as following:

Problem 1.2.1. *Given an optimal control problem, find the quantization function* q*,*

$$q : \mathcal{D} \subset M \rightarrow \mathcal{Q},$$

where $\mathcal{D}$ *is a domain of the state manifold* M *and* $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ *is a finite set of regions of the state space* M *such that* $\cup Q_i = \mathcal{D}$ *and a mapping*

$$p \circ q : \mathcal{D} \subset M \rightarrow U$$

such that the trajectories derived from the quantized control system

$$\dot{x} = f(x, p \circ q(x))$$

are ε*-close to the continuous optimal trajectories of the given optimal control problem.*

Current numerical techniques for the Dynamic Programming method cannot be implemented to problems of high dimensions due to the "curse of dimensionality". Hence numerical computation of feedback control laws seems to be a far-fetched proposition at the moment. We need more information on the geometric structure of our optimal control problem. The hope is that it would be possible to derive feedback controls from geometrical objects like curvature, torsion, etc which are measures of deviation from flat Euclidean spaces. Physics has especially benefited from the study of curved manifolds through Riemannian geometry. However, optimal control problems do not lie on Riemannian manifolds but on more general spaces called Finsler spaces. Recently there has been an abundance of literature in geometric optimal control theory, see [2],[16],[39],[54], however they mostly take the symplectic view and not that of Finsler geometry. Studying the exact transformation of optimal control problems into Finsler spaces and deriving the geometrical objects for numerical computation is the main goal of this thesis.

The other objective of this thesis is to come up with a symbolic algorithm as much as possible keeping in view the design and modelling aspects of control systems. Most numerical methods for optimal control problems do not take the design view at all into consideration. They simply compute optimal controls of a given control system without any verification that the controls generated are indeed optimal. What is also required in real-life situations is an analysis of the control system upon change in design parameters and the variations in the model itself. Hence symbolic algorithms make a strong case for design and modelling of control systems. Of course we note that symbolic algorithms alone cannot solve the control problems entirely and hence numerical methods will eventually come into play.

1.3 Thesis Contribution

Optimal control in a coordinate free manner on manifolds has been carried out by several authors, notably Sussmann [67], [68], who stated the Pontryagin Maximum Principle in a coordinate independent notation. There the variational and the adjoint equations of the maximum principle are exactly the parallel transport equations of vectors and covectors, respectively, of local connections defined as Lie brackets. Another method mentioned in that literature is the iterated Hamiltonian lift, which has a symplectic flavor. Finsler methods for formulating optimal control have been suggested by several people, including [49], [71]. Lopez [49], suggests that indicatrices are the primary objects of study, possibly because bang-bang controls appear as holes on the indicatrices. But this does not yield computation techniques. Udriste [71] proposed using the Okubo technique(see [4], [8], [28]) for generating metrics. None of these investigations has yet led to algorithms for computing optimal controls, but then coordinate free methods are more for beauty, understanding, generality than for computation.

In 1992-94 Kohn and Nerode observed that every optimal control problem gives rise to a Finsler metric ground form on a suitable manifold. They began the development of optimal control theory based on Finsler Geometry in [43], [44], [45] with the goal in mind of computing near optimal controls in real time. Optimal trajectories are Finsler geodesics, optimal controls are Finsler connections. The principal problem is to compute them, preferably by real time algorithms. However, the Finsler metric derived here is for the Lagrangian with fixed control and hence they were only solving calculus of variations problems and not that of optimal control problems. In order to deal with optimal control problems, the metric has to simultaneously take into consideration, the state, the controls, and the adjoint variables. The development of a metric for optimal control problems is the main contribution of this thesis.

This thesis incidentally establishes the exact relations between the Finsler geodesic formulation and the Pontryagin maximum principle. We explicitly derive the geodesic equations in the presence of state equality constraints and further develop symbolic and numerical computation tools for determining connections and curvature terms for the eventual purpose of computing close to optimal control for real systems

Connections, covariant derivatives, parallel transport, are equivalent notions which were introduced into the geometry of manifolds by Ricci and his student Levi Civita in about 1900 to give meaning to the notion that a tangent line at one point on a curve is parallel to a tangent line at another point on that same curve. There can be many connections on a manifold. The notion cannot be defined in terms of intrinsic manifold structure alone or in terms of a metric ground form such as in Riemannian geometry, where the metric ground form is positive definite, or in terms of the metric ground form $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 - dt^2$ of Einstein's four dimensional space-time. Einstein used a connection defined by a "covariant differentiation" to transfer notions of length and angle from one point to another along space time trajectories. In the 1970's, connections became ubiquitous in physics as the proper expression of the notion of physical field. The current abstract coordinate-free definition was given by Kozul in the 1950's(see [51]).

The oldest example, from Levi Civita, is the Euclidean connection on an open subset U in Euclidean space $\mathbb{R}^n$, is defined as a "covariant derivative"

$$D_u V\Big|_x = \left(u^i \frac{\partial V^1}{\partial x^i}, \ldots, u^i \frac{\partial V^n}{\partial x^i} \right),$$

where $V = (V^1, \ldots, V^n) : \mathcal{U} \to \mathbb{R}^n$ is a C^∞ vector field on $\mathcal{U}$ and $u = (u^1, \ldots, u^n) \in T_x\mathcal{U} = \mathbb{R}^n$ is a tangent vector at $x \in \mathcal{U}$.

The Euclidean connection for a manifold embedded in Euclidean space may be thought of as dragging a tangent vector parallel to itself along the curve. But this description is dependent on the embedding of the manifold in Euclidean space. To get an intrinsic definition, Levi Civita looked at parallel transport by the connection of the coordinate lines of a Euclidean coordinate system at a point on a curve to an infinitisimally close point of the curve. It is an infinitisimal rotation. Integrating the equations for that rotation gave the equations of parallel transport. They describe how Euclidean coordinates change as one moves along a curve by a semigroup of invertible linear transformations along the curve. The covariant derivative can be so described.

In studying the change of the vector field $V \in C^\infty(TM)$ along a given curve C we have to consider two factors as we pass from p to q, where $p, q \in M$. First,: the change $dV^i = \frac{dV^i}{dt}\, dt$ in $V^i(t)$ which depends solely on the definition of the vector field V^i and is independent of the metric of the space. Second, the change in metric from tangent space T_pM to tangent space T_qM. In the optimal control, the variational equation and the adjoint equation of the Pontryagin Maximum Principle are parallel transport equations for vectors and covectors, respectively. In this thesis, we derive the Finsler versions which allow us to transport controls along state trajectories.

In order to derive control laws in presence of disturbances one needs to derive controls as a function of state, $u(x)$, which are usually called feedback controls. The synthesis problem for optimal controls remains a principal problem. However, if one assumes that the disturbances in the system dynamics are small in magnitude and bounded then one should be able to derive those optimal controls as a function of those disturbances of state. In this thesis, we deal with this problem of synthesizing controls which takes into account the geometry of the given optimal control problem. We do not synthesize for the whole state space(globally) but only for trajectories close(locally) to the optimal trajectories.

The derivation a Finsler metric for a given optimal control problem and its corresponding affine connection, gives us a way of differentiating vector fields on this manifold along a given direction. We show how this connection gives us certain geometric properties of the manifold in terms of torsion and curvature which are third and fourth order tensors respectively.

On Riemannian manifolds there exists an unique affine (Levi-Civita)connection which is torsion free and metric compatible. The curvature tensor which is an intrinsic object of the manifold helps in studying geodesic flows. In Euclidean spaces the curvature term always vanishes since the derivatives $\partial_i \partial_j$ always commute, i.e.,

$$\frac{\partial^2}{\partial x^i \partial x^j} = \frac{\partial^2}{\partial x^j \partial x^i}.$$

However this is not the case on Riemannian manifolds, as the commutator of covariant derivatives defined by the Levi-Civita connection, $[D_i, D_j]$ does not vanish and is given by the following index free notation

$$R(U,V)W = (D_U D_V - D_V D_U - D_{[U,V]})W, \tag{1.9}$$

where U, V, and W are vector fields on M.

In Finsler geometry the curvature tensor not only depends on the position but also the direction. Here we derive the curvature tensors from the Chern connection using Cartan's structural equation. We do not have just one curvature but have two different objects known in literature as Riemann and Chern curvatures. The Riemann curvature gives us a measure of the the dispersion of the geodesics with different initial conditions. The Chern curvature can be derived from Cartan's torsion, which is defined in Minkowski space. Since every tangent space on a Finsler space is a Minkowski space, Cartan torsion is defined for a Finsler space. We outline the derivation of these two curvature tensors in the thesis and study how they affect the spread of the optimal controls for neighboring points on the manifold.

The other main contribution of this thesis is the utilization of the geometric objects we have derived for a given optimal control problem to compute optimal controls.

In this regard the Taylor expansion method is used as the numerical integrators for the geodesic equations. This leads to expressing them in terms of Riemannian curvature and the geodesic deviation. Using curvature terms in the computation of controls and errors is a novelty of this thesis.

Typical numerical methods for optimal control problems can be classified into direct and indirect methods. There exists many good surveys and books for these methods in the literature, see [14],[60],[70]. Indirect methods use the necessary conditions to compute optimal controls. Usually if the switching structure is known apriori then the indirect methods lead to a greater degree of accuracy. They focus on obtaining a solution to the classical necessary conditions for optimality which take the form of a two-point boundary value problem. One of the common variations on the indirect method is the steepest descent algorithm. In this method, the state equation is integrated forward using a guess for the control profile, and then the costate equation is integrated backward. The maximality condition is then used locally to find a steepest descent direction for u at a discrete number of points, and globally as a termination criterion. Backward integration of the costate equations may be avoided by using a direct shooting method (see [14],[15],[69]), in which initial values of the adjoint variables are guessed initially, and updated iteratively. There are some problems with this method, including a small region of convergence to the optimum, the need to formulate the costate equations, and the difficulties in choosing reasonable initial guesses for the adjoint variables. The multiple shooting method (see [40, 62]) was proposed to extend the small convergence region of the direct shooting method. Multiple shooting algorithms transform the problem into a multi-point boundary value problem for the state and adjoint variables. Homotopy methods [72] have also been proposed as a response to the small region of convergence. Other indirect techniques have been the finite-difference method [46] which is essentially a low-order collocation method, and standard collocation techniques [6].

On the other hand, direct methods are a result of discretization of the original problem and recasting as a finite-dimensional optimization problem, typically a nonlinear program (NLP), see [42]. They are used especially for tackling large-scale problems by using standard techniques from nonlinear programming. There are two general strategies within the framework of the direct method: sequential method and simultaneous strategy methods. In the sequential method often called control parametrization, the control variables are discretized over finite elements using polynomials or in fact any suitable basis functions (see [14],[15],[70]). The coefficients of the polynomials and the size of the finite elements then become decision variables in a master nonlinear program. Function evaluation is carried out by solution of an initial value problem of the original dynamic system, and gradients for a gradient-based search may be evaluated by solving either the adjoint equations or the sensitivity equations. In the simultaneous strategy both the controls and the state variables are discretized using polynomials on finite elements, and the coefficients and elements sizes become decision variables in a much larger NLP (see [75]). Unlike control parametrization, the simultaneous method does not require the solution of initial value problems at every iteration of the NLP.

There are other methods like the ε-method by Balakrishnan [7] where the dynamic constraints are entirely removed by a combination of penalty function methods and Rayleigh-Ritz type expansion methods. It has been shown that the optimal values of the penalized problem converge to the optimal value of the relaxed problem. Techniques such as gradient methods in function spaces using penalty function methods and feasible directions methods can be found in [55],[56],[60]. Iterative methods which are essentially based on solutions of the Riccati differential equations appearing in linear quadratic optimal control problems are also used in computing optimal controls, see [41], [37]. Such methods solve the original nonlinear problem through a sequence of linear quadratic optimal control problems which are suitably formulated as the accessory optimal control problems to the original one. This method exhibits fast convergence properties. However for large scale problems it is avoided due to derivative computations.

Finally, the Hamilton-Jacobi-Bellman equation which is derived from Bellman's principal of optimality has its own set of techniques. However this is a partial differential equation and in practice it is very difficult to solve except in certain fortuitous cases, see [10], [9]. There has been a lot of research in computing optimal controls in the aerospace, chemical processes, and recently finance communities.

Unlike the above mentioned algorithms, we start with an initially known control law, which transports the system from the current state to the desired goal in an optimal fashion. Our algorithm begins with a subdivision of the total cost V^0 of the initial path into N intervals with nodes

$$0 < s_1 < s_2 < \cdots s_{N-1} < s_N = V^0.$$

The algorithm first finds a geodesic between the state nodes $x_1^{(0)}$ and $x_3^{(0)}$ which could be done by the simple shooting method. The next step involves transporting the midpoint $x_2^{(0)}$ of the initial trajectory to the new geodesic and rename it $x_2^{(1)}$. This process is continued by obtaining the geodesic between nodes $x_2^{(1)}$ and $x_4^{(0)}$ and choosing the midpoint as $x_3^{(1)}$. In this way we get a new trajectory $x^{(i)}(s)$ after each iteration. The numerical integrators for the geodesics are in turn developed using the curvature terms for the given problem. Furthermore, since the geodesic equations derived from the Finsler metric for a given optimal control problem consist of differential equations for the state, control, and the adjoint variables simultaneously, and with the knowledge of the local costs along the geodesics between local nodes, we could introduce better adaptive techniques which other algorithms do not consider.

The thesis follows with a chapter introducing the calculus of variations and optimal control and tieing them up together by pointing out how optimal control problems are a generalization of the calculus of variations problems. It also details the existing literature on lifting the optimal control problems to geometry. Chapter

3 devises a Finsler metric for optimal control problems and derives the necessary conditions for a control function to be optimal. It also derives necessary conditions for several sample control vector fields. Chapter 4 introduces control sprays and defines Finsler metrizability for control sprays. Chapter 5 derives the neighboring optimal controls for control systems in the Finsler setting which involves the use of Riemann curvature terms. Chapter 6 presents numerical methods to compute optimal controls using the curvature terms. Finally, Chapter 7 concludes with my future research plans and further questions instantiated by this thesis.

Chapter 2

Optimal Control, Calculus of Variations, and Geodesics

2.1 Calculus of Variations

The calculus of variations is a 300 year old subject, developed by many famous mathematicians in many directions. Optimal control is a 50 year old subject developed through the efforts of Pontryagin and his successors. Either subject can be formulated in terms of the other, although the exact problems experts in each field have studied are not identical. Because of the vast variational literature, it is desirable to be able to shift rapidly from one point of view to the other. In this chapter we explain why the Pontryagin maximum principle, a necessary condition for the controls to be optimal, is roughly equivalent to the conjunction of the usual necessary conditions of the calculus of variations taken, namely Euler-Lagrange, Hamilton-Jacobi, and the Weierstrass sufficiency. Finsler geometry gives yet a third language for these problems. We deal with this separately. Different formulations make different symbolic formulas and algorithms float to the surface.

The calculus of variations deals mainly with optimization problems involving a function, usually called the Lagrangian,

$$(x, \dot{x}) \mapsto L(x, \dot{x}, t)$$

of "position, velocity, and time". The position or state x is usually taken to be a vector $x = (x^1, \ldots, x^n)$ in n-dimensional Euclidean space $\mathbb{R}^n$, and the velocity $\dot{x} = (\dot{x}^1, \ldots, \dot{x}^n)$ is then also a vector in $\mathbb{R}^n$.

The objective of a typical calculus of variations problem is to find a curve $t \mapsto \xi(t)$ in x-space that minimizes the integral

$$I(\xi) = \int_a^b L(\xi(t), \dot{\xi}(t), t)dt$$

of the Lagrangian, among all curves in Ξ, where

$$\Xi := \{\xi : [a, b] \to x(t) \in \Omega \mid \xi(a) = x_a \text{ and } \xi(b) = x_b\},$$

where Ω and open region in $\mathbb{R}^n$ is called the state space (set of all possible values of x), and for the time being can be considered to be an open subset of n-dimensional

Euclidean space $\mathbb{R}^n$. Let us now make the velocity vector $\dot{x}$ an independent variable by renaming it as u which makes the Lagrangian L as a function of three independent variables (x, u, t), i.e.,

$$L : \Omega \times \mathbb{R}^n \times [a, b] \to \mathbb{R}.$$

2.1.1 Euler-Lagrange equation

We now look for necessary conditions for a given curve $t \mapsto \xi_*(t)$ to be a solution to the calculus of variations problem. The general condition known today as the "Euler-Lagrange equation" was derived by Leonhard Euler and Joseph-Louis Lagrange. The Euler-Lagrange equation says that the identity

$$E(x, x', t) \ := \ \frac{d}{dt}\frac{\partial L}{\partial u^i} - \frac{\partial L}{\partial x^i} = 0, \qquad i = 1, \ldots, n \tag{2.1}$$

must be satisfied along the curve ξ_*. This necessary condition for an optimum is derived by looking at the first variation δI of the cost functional I, and requiring that $\delta I = 0$, i.e., by linearizing the inequality

$$\delta I(u_0) = \int_a^b L(t, x_0 + \delta x, (x_0 + \delta x)') - L(t, x_0, x_0') \geq 0. \tag{2.2}$$

with respect to an infinitesimal small and localized variation

$$\delta x = \begin{cases} \rho(t) & \text{if } \ t \in [t_0, t_0 + \varepsilon], \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

Here $\rho(t)$ is a smooth function that vanishes at points t_0 and $t_0 + \varepsilon$ and is constrained as follows:

$$|\rho(t)| < \varepsilon, \qquad |\rho'(t)| < \varepsilon, \qquad \forall t \in [t_0, t_0 + \varepsilon].$$

Linearizing and integration by parts of inequality (2.2), we have

$$\delta I(x_0) = \varepsilon \int_a^b E(x, x', t)\delta x + \frac{\partial L}{\partial x'}\delta x\Big|_{t=a}^{t=b} + o(\varepsilon) \geq 0. \tag{2.4}$$

The second term vanishes because the boundary values of x are given and their variations are $\delta x|_{t=a} = \delta x|_{t=b} = 0$

Theorem 2.1.1. Assume n is a positive integer, Ω is an open subset of $\mathbb{R}^n$, a, b are real numbers such that $a < b$, L is a real-valued function on $\Omega \times \mathbb{R}^n \times [a, b]$, and x_a and x_b are given in Ω. Assume that L is a function of class C^2. Let $\xi_* : [a, b] \to \Omega$ be a curve of class C^1 which is a solution of the minimization problem $P(L, a, b, x_a, x_b)$ in the space of all curves of class C^1. Then the Euler-Lagrange equation

$$E(x, \dot{x}, t) = 0$$

holds for all $t \in [a, b]$.

In addition to being a solution to the Euler-Lagrange equation, the true minimizer satisfies necessary conditions in the form of inequalities. These conditions distinguish the trajectories that correspond to the minimum of the functional from trajectories that either correspond to its maximum or to a saddle point stationary solution. We now look at two such conditions which help us to understand the maximum principle.

2.1.2 The Legendre condition

The next natural step is to look at the second variation of I, and this was done by *Adrien-Marie Legendre*, who found an additional necessary condition for a minimum which states that the Hessian matrix $\{\frac{\partial^2 L}{\partial u^i \partial u^j}\}_{1\leq i,j\leq n}$ has to be nonnegative definite, i.e.,

$$\sum_{i,j=1}^{n} \alpha^i \alpha^j \frac{\partial^2 L}{\partial u^i \partial u^j}(\xi_*(t), \dot{\xi}_*(t), t) \geq 0 \qquad \text{for all} \quad \alpha = (\alpha^1, \dots, \alpha^n) \in \mathbb{R}^n. \tag{2.5}$$

2.1.3 Hamilton's form

Suppose a curve $t \mapsto \xi_*(t)$ is a solution of the Euler-Lagrange equation. Define a function $(x, u, p, t) \mapsto H(x, u, p, t)$ of three vector variables x, u, p in $\mathbb{R}^n$, and of $t \in \mathbb{R}$, by letting

$$H(x, u, p, t) = \langle p, u \rangle + L(x, u, t), \tag{2.6}$$

and call it the control Hamiltonian. Here we depart from the Hamiltonian that Hamilton actually defined. Then if we define the momentum, $\zeta(.) : [a, b] \to p(t) \in \mathbb{R}^n$ by

$$\hat{\zeta}(t) = -\frac{\partial L}{\partial u}(\xi_*(t), \dot{\xi}_*(t), t). \tag{2.7}$$

Differentiating the control Hamiltonian (2.6) with respect to the momentum variable p gives us

$$\frac{\partial H}{\partial p} = u,$$

from which it follows that along the curve ξ_*:

$$\dot{\xi}_*(t) = \frac{\partial H}{\partial p}(\xi_*(t), \dot{\xi}_*(t), p(t), t). \tag{2.8}$$

Similarly, differentiating the control Hamiltonian with respect to the state vector x gives

$$\frac{\partial H}{\partial x} = -\frac{\partial L}{\partial x}. \tag{2.9}$$

From the Euler-lagrange equation it then follows that

$$\dot{\zeta}(t) + \frac{\partial H}{\partial x}(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t) = 0. \tag{2.10}$$

Finally, differentiating with respect to the control u gives

$$\frac{\partial H}{\partial u} = p + \frac{\partial L}{\partial u}, \tag{2.11}$$

and since we have defined the momentum vector p by equation (2.7) we have

$$\frac{\partial H}{\partial u}(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t) = 0. \tag{2.12}$$

So we have shown that the Euler-Lagrange equation (2.1) can be written in the form

$$\frac{dx}{dt} = \frac{\partial H}{\partial p}, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial x}, \qquad \frac{\partial H}{\partial u} = 0. \tag{2.13}$$

We now show what Hamilton actually wrote down as the Hamiltonian system and how it is different from the control Hamiltonian which leads us to the maximum principle. We begin with Hamilton's hamiltonian $\mathcal{H}$ which is a function $(x, p, t) \mapsto \mathcal{H}(x, p, t)$, defined by the formula

$$\mathcal{H}(x, p, t) = \langle p, \dot{x} \rangle + L(x, \dot{x}, t), \tag{2.14}$$

which is similar to the control Hamiltonian but differs in that the velocity vector $\dot{x}$ is considered to be a dependent variable and is defined implicitly by the equation

$$p = \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t). \tag{2.15}$$

This leads to the Hamiltonian system of equations

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p}, \qquad \frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial x}, \tag{2.16}$$

which is equivalent to the system if the map $(x, \dot{x}, t) \mapsto (x, p, t)$ defined by equation (2.15) can be inverted, i.e., if we can solve for $\dot{x}$ as a function of x, p, t. So the classical Hamiltonian system only makes sense when the transformation can be inverted, at least locally, while the control Hamiltonian is equivalent to the Euler-Lagrange system under more general conditions. This transformation is called traditionally the Legendre transform.

Furthermore, since the control Hamiltonian is equal to $L(x, u, t)$ plus a linear function of u, its clear that the Legendre's condition is equivalent to

$$\frac{\partial^2 H}{\partial u^2}(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t) \leq 0 \tag{2.17}$$

Now if we rewrite the equations (2.12) and (2.17) together

$$\frac{\partial H}{\partial u} = 0, \qquad \text{and} \qquad \frac{\partial^2 H}{\partial u^2} \leq 0, \tag{2.18}$$

it is immediately clear that the Euler-Lagrange and Legendre conditions are on the critical points with respect to the velocity variable u of the control Hamiltonian H. One can see that the Euler-Lagrange and Legendre conditions are first and second order conditions about critical points of the same function H. The critical point conditions of the Lagrangian L and the Hamilton's $\mathcal{H}$ do not satisfy the Euler-Lagrange and Legendre conditions.

2.1.4 Weierstrass excess function

We now look at the test which detects stability of a solution to a variational problem against strong local perturbations. Weierstrass introduced what is known as the *Weierstrass excess function*,

$$\mathcal{E}(x, \bar{u}, u) = L(x, u) - \frac{\partial L}{\partial u}(x, \bar{u}) \cdot u. \tag{2.19}$$

This function depends on three sets of independent variables, namely, x, $\bar{u}$, and u, i.e., $\mathcal{E}$ is a function of position and two velocity vectors. Weierstrass then proved his side condition which states that: *for a curve $s \mapsto \xi_*(S)$ to be a solution of the minimization problem, the excess function $\mathcal{E}$ has to be ≥ 0 when evaluated for $x = \xi_*(s)$, $\bar{u} = \dot{\xi}_*(s)$, and a completely arbitrary u.* By the homogeneity property of the Lagrangian in the velocity vector, we have

$$L(x, u) = \frac{\partial L}{\partial u}(x, u) \cdot u. \tag{2.20}$$

Therefore, the excess function could now be written as

$$\mathcal{E}(x, \bar{u}, u) = L(x, u) - \frac{\partial L}{\partial u}(x, \bar{u}) \cdot u - (L(x, \bar{u}) - \frac{\partial L}{\partial u}(x, \bar{u}) \cdot \bar{u}). \tag{2.21}$$

The variation used in the Weierstrass test is an infinitesimal triangle supported on the interval $[t_0, t_0 + \varepsilon]$ in a neighborhood of a point $t_0 \in [a, b]$:

$$\delta x(t) \;=\; \begin{cases} 0, & \text{if } \; t \notin [t_0, t_0 + \varepsilon], \\ v_1(t - t_0), & \text{if } \; t \in [t_0, t_0 + \varepsilon], \\ v_1 \alpha\varepsilon + v_2(t - t_0 - \alpha\varepsilon), & \text{if } \; t \in [t_0, t_0 + \varepsilon], \end{cases} \tag{2.22}$$

where the parameters α, v_1, and v_2 are related by

$$\alpha v_1 + (1 - \alpha) v_2 = 0.$$

The Weierstrass variation is localized and has an infinitesimal absolute value if $\varepsilon \to 0$, but its derivative $\delta x'$ is finite, unlike the variation in (2.3):

$$\delta x' \;=\; \begin{cases} 0, & \text{if } \; t \notin [t_0, t_0 + \varepsilon], \\ v_1, & \text{if } \; t \in [t_0, t_0 + \varepsilon], \\ v_2, & \text{if } \; t \in [t_0, t_0 + \varepsilon]. \end{cases} \tag{2.23}$$

Using the momentum variable $\zeta(t)$ where

$$\zeta(t) = \frac{\partial L}{\partial u}(\xi_*(t), \dot{\xi}_*(t), t), \tag{2.24}$$

we see that

$$\mathcal{E}(\xi_*(t), \dot{\xi}_*(t), u, t) = (L(\xi_*(t), u, t) - \langle \zeta(t), u \rangle) - (L(\xi_*(t), \dot{\xi}_*(t), t) - \langle \zeta(t), \dot{\xi}_*(t) \rangle), \tag{2.25}$$

which could alternately be rewritten using the control Hamiltonian H as

$$\mathcal{E}(\xi_*(t), \dot{\xi}_*(t), u, t) = H(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t) - H(\xi_*(t), u, \zeta(t), t). \tag{2.26}$$

So the Weierstrass condition in control Hamiltonian form says that: along an optimal curve $t \mapsto \xi_*(t)$, if we define $\zeta(t)$ by equation (2.24)then for every t the value $u = \dot{\xi}_*(t)$ must maximize the Hamiltonian $H(\xi_*(t), u, \zeta(t), t)$ as a function of t. Furthermore, we do not need to define $\zeta(t)$ by (2.24), as maximizing the control Hamiltonian with respect to u gives us a first order condition which is equivalent to (2.24). So we drop (2.24) altogether and rewrite the optimality condition as if a curve $t \mapsto \xi_*(t)$ is a solution of the minimization problem. Then there has to exist a function $t \mapsto \zeta(t)$ such that the following three conditions hold for all t:

$$\begin{aligned} \dot{\xi}_*(t) &= \frac{\partial H}{\partial p}(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t), \\ \dot{\zeta}(t) &= -\frac{\partial H}{\partial x}(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t), \\ H(\xi_*(t), \dot{\xi}_*(t), \zeta(t), t) &= \max_u H(\xi_*(t), u, \zeta(t), t). \end{aligned}$$

The above conditions can be viewed as the Pontryagin maximum principle for the calculus of variations problem. It quite neatly simplifies three different necessary conditions of Euler-Lagrange, Hamilton, and Weierstrass into just one set of conditions.

2.1.5 Nonconvex Variational Problems : Relaxation

In most difficult applications, the Lagrangian $L(t, x, x')$ is nonconvex with respect to the velocity variable x'. In this case, the Weierstrass test fails, and the problem becomes ill-posed. Suppose the Lagrangian $L(t, x, x')$ is bounded from below (say, by zero),

$$L(t, x, x') \geq 0, \qquad \forall t, x, x';$$

and satisfies the condition

$$\lim_{|x'| \to \infty} \frac{L(t, x, x')}{|x'|} = \infty.$$

Then the infimum I_0

$$I_0 = \inf_x I(x), \qquad I(x) = \int_a^b L(t, x, x')dt \tag{2.27}$$

is nonnegative, $I_0 \geq 0$.

There is of course a minimizing sequence $\{x^i\}$ such that $I(x^i) \to I_0$. However, since $L(., ., x')$ is not convex, this minimizing sequence will not in general converge to a piecewise smooth curve in the limit. Otherwise it would have satisfied the Euler-Lagrange equation and the Weierstrass test, but the last requires convexity of $L(., ., x')$. But such a minimizing sequence converges to a "generalized curve" which consists intuitively of infinitesimal zigzags. The limiting curve can have a derivative with a dense set of points of discontinuity. A detailed explanation of this phenomenon can be found in Young [76] and Warga [73]. The existence theorem involved goes back to McShane 1939-40 in the calculus of variations.

Example 3. Consider a simple variational problem that yields to the generalized solution:

$$\inf_x I(x) = \inf_x \int_0^1 G(x, x')dt,$$

where

$$G(x, z) = x^2 + \min\{(z-1)^2, (z+1)^2\}, \qquad x(0) = x(1) = 0.$$

(Intuition: The Lagrangian G attaches a penalty to the trajectory x when it has speed $|x'|$ different from ± 1. It also attaches a penalty to the trajectory for any deflection of the trajectory x from zero. These contradictory requirements cannot be met by a piecewise smooth trajectory.)

A minimizing sequence for the above problem can be constructed in the following way. Consider a set of functions $\tilde{x}(t)$ that belong to the boundary of the forbidden interval $\tilde{x}'(t) = 1$ or $\tilde{x}'(t) = -1$ of the nonconvexity of $G(., z)$. The second term in the Lagrangian vanishes when these function are substituted, and the problem becomes

$$I(\tilde{x}, \tilde{x}') = \int_0^1 \tilde{x}^2 dt.$$

(Observation: the term $\tilde{x}$ oscillates near zero if the derivative $\tilde{x}'$ changes its sign on intervals of equal length. The cost $I(\tilde{x}, \tilde{x}')$ depends on the density of switching points and tends to zero when the number of these points increases. Therefore a minimizing sequence can be defined as a sequence of saw-tooth functions $\tilde{x}$. The the minimizing sequence $\{\tilde{x}^i\}$ does not converge to any piecewise smooth function,

but rather to a measure valued function. For nonconvex problems, the classical variational technique based on the Euler-Lagrange equations in their usual form fails to work. But optima exist as weak solutions which are measure valued in a suitable completion of the space of ordinary solutions .)

Convex Envelope

Consider a variational problem with a nonconvex Lagrangian L. We shall replace this problem with a new one which describes infinitely rapidly oscillating minimizers in terms of averages. This will be done by constructing the convex envelope of nonconvex Lagrangian.

Definition 2.1.2. *The convex envelope* $\mathsf{co}\, L$ *is a solution of the following minimal problem:*

$$\mathsf{co}\, L(v) \;=\; \inf_{\xi} \frac{1}{b-a} \int_a^b L(v+\xi), \qquad \forall \xi : \int_a^b \xi = 0. \tag{2.28}$$

This definition determines the convex envelope as the minimum of all parallel secant hyperplanes that intersect the graph of L. To compute the convex envelope $\mathsf{co}\, L$, one can use the Caratheodory's theorem. This theorem allow us to replace the integral in the right-hand side of the definition of $\mathsf{co}\, L$ by the sum of $n+1$ points which leads to the following optimization problem:

$$\mathsf{co}\, L \;=\; \min_{m_i \in M} \min_{\xi_i \in \mathcal{E}} \sum_{i=1}^{n+1} m_i L(v+\xi_i), \tag{2.29}$$

where

$$M = \left\{ m_i \;:\; m_i \geq 0, \;\; \sum_{i=1}^{n+1} m_i = 1 \right\} \tag{2.30}$$

and

$$\mathcal{E} = \left\{ \xi_i \;:\; \sum_{i=1}^{n+1} m_i \xi_i = 0 \right\}. \tag{2.31}$$

The convex envelope $\mathsf{co}\, L(v)$ of a function $L(v)$ at a point v coincides with either the function $L(v)$ or the hyperplane that touches the graph of the function L. The hyperplane remains below the graph of l except at the tangent points where they coincide. The convex envelope could also be defined as the greatest convex function that does not exceed $L(v)$ at any point v:

$$\mathsf{co}\, L(v) = \max \phi(v) \;\;:\;\; \forall v, \; \phi(v) \leq L(v), \text{ and } \phi(v) \text{ is convex.}$$

Relaxation

To deal with a nonconvex problem, we "relax" it. Relaxation means that we replace the problem with another one that has the same cost but whose solution is stable against fine-scale perturbations; particularly, it cannot be improved by the Weierstrass variation. The relaxed problem has a classical solution and the infimum of the functional in the initial problem coincides with the cost of this classical solution.

Consider the extremal problem (2.27) and the corresponding solution $x_0(t)$. Let us fix two neighboring points $A = (t_0, x_0(t_0))$ and $B = (t_0 + \varepsilon, x_0(t_0 + \epsilon))$ on this solution. Using Taylor expansion we can represent the point B as

$$B = (t_0 + \varepsilon, x_0(t_0) + \varepsilon x'(t_0) + o(\varepsilon)).$$

The new cost due to this interval could be written as

$$I_\varepsilon(x_0) = \int_{t_0}^{t_0+\varepsilon} L(t, x_0, x_0') = \varepsilon L(t, x_0(t_0), x_0'(t_0)) + o(\varepsilon).$$

Let us examine a local variation of the solution $x_0(t)$ by replacing it with a zigzag piecewise linear curve x_ε that passes through the points A and B. Consider a continuous curve x_ε that contains $p-1$ subintervals of the constancy of the derivative $v = x_\varepsilon'$. The derivative x_ε' is given by

$$x_\varepsilon'(t) \ = \ x_0'(t_0) + v_k, \qquad \text{if} \quad t \in [t_0 + \varepsilon \sum_{i=1}^{k} m_i, t_0 + \varepsilon \sum_{i=1}^{k+1} m_i], \tag{2.32}$$

where $k = 1, \dots, p-1$. The saw-tooth curve x_ε is then

$$x_\varepsilon(t) \ = \ x_0(t_0) + \int_{t_0}^{t} (x_0'(t_0) + v(t))dt. \tag{2.33}$$

We require that any admissible solution x_ε passes through point B, i.e., its value at the point $t_0 + \varepsilon$ must be equal to $x_0(t_0 + \varepsilon)$ up to the terms of the order of $o(\varepsilon)$,

$$x_\varepsilon(t_0 + \varepsilon) - x_0(t_0 + \varepsilon) = \sum_{i=1}^{p} m_i v_i = o(\varepsilon).$$

Let us now estimate the integral of $L(t, x_\varepsilon, x_\varepsilon')$ over this interval, up to the order of $o(\varepsilon)$:

$$L(t, x_\varepsilon(t), v_i + x_0'(t_0)) = L(t_0, x_0(t_0), x_0'(t_0) + x_\varepsilon'(t)) + o(\varepsilon),$$

for any $t \in [t_0, t_0+\varepsilon]$. The Lagrangian is piecewise constant in the interval $[t_0, t_0+\varepsilon]$. The impact of $I_\varepsilon(x_\varepsilon)$ becomes

$$I_\varepsilon(x_\varepsilon) = \varepsilon \sum_{i=1}^{p} m_i L(t_0, x_\varepsilon(t_0), x_0'(t_0) + v_i) + o(\varepsilon). \tag{2.34}$$

Calculate the minimum of (2.34) with respect to the arguments $v_1, \ldots, v_p$ and $m_1, \ldots, m_p$, which are subject to the constraints

$$m_i(t) \geq 0, \quad \sum_{i=1}^{p} m_i = 1, \quad \sum_{i=1}^{p} m_i v_i = 0. \tag{2.35}$$

This minimum coincides with the convex envelope of the original Lagrangian with respect to its velocity argument and moreover due to Caratheodary's theorem, it is enough to split the interval into $p = n + 1$ parts.

Comparing the costs $I_\varepsilon(x_0)$ and $I_\varepsilon(x_\varepsilon)$ corresponding to the smooth solution x_0 and to the zigzag solution x_ε, we obtain the following inequality:

$$\frac{1}{\varepsilon}(I_\varepsilon(x_0) - I_\varepsilon(x_\varepsilon)) \;=\; L(t_0, x_0(t_0), x_0'(t_0)) - \mathsf{co}\; L(t_0, x_\varepsilon(t_0), x_\varepsilon'(t_0)) + o(\varepsilon) \;\geq 0.$$

We see that the zigzag solution x_ε corresponds to lower cost in the regions of non-convexity. Passing to the variational problem in the whole interval $[0, 1]$ we can rewrite the relaxed problem as

$$I = \min_x \int_0^1 \mathsf{co}\; L(t, x(t), x'(t)) dt.$$

The curve x_ε converges strongly to the curve x_0 in the sense that:

$$\|x_\varepsilon - x_0\|_{L_\infty[0,1]} \to 0, \qquad \text{as } \varepsilon \to 0,$$

but its converges to x_0' only weakly in L_p,

$$\int_0^1 \phi(x_\varepsilon') - x_0') \to 0, \quad \forall \phi \in L_q[0,1], \;\; \frac{1}{p} + \frac{1}{q} = 1.$$

This pointwise convergence for all ϕ in a function space is weak convergence. It is useful for proving the existence of optimal controls in more general contexts. Approximations in this convergence can be used as implementable controls with near optimal performance. A good reference for the analysis is Shilov [66].

2.2 Optimal Control

2.2.1 Control System

In optimal control theory we are given a system whose dynamics follow the following under-determined system of ordinary differential equations:

$$\dot{x} = f(x,u), \quad \text{for } x \in M \subseteq \mathbb{R}^n, \;\; u \in U \subseteq \mathbb{R}^m, \tag{2.36}$$

where $x = (x^i)$ represents the state space variables, $u = (u^i)$ represents the control variables and $f : M \times U \to TM$ is a function. We assume that, for each $x_0 \in \mathbb{R}^n$, the map $u \mapsto f(x_0, u)$ is an embedding of $\mathbb{R}^m$ into $\mathbb{R}^n$, i.e., the controls should appear non-degenerately in the equations.

From the geometric point of view, it can be understood as a fibred mapping

$$X : \mathbb{U} \to M \tag{2.37}$$

from a control fibre bundle $(\mathbb{U}, \kappa, M)$ over the state manifold M to the tangent bundle (TM, π, M). Here the fibres of the control fibre bundle $(\mathbb{U}, \kappa, M)$, $\kappa^{-1}(x)$, $x \in M$ denote the state dependent input spaces. The above geometric description can be best described by the following commutative diagram:

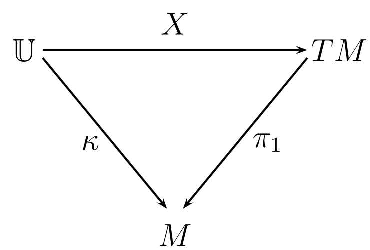

Figure 2.1: Control System

In local coordinates (x^i) in M, adapted coordinates (x^i, u^j) in U, and natural coordinates (x^i, v^i) in TM, the coordinate expression for X is

$$X(x,u) = f^i(x,u)\frac{\partial}{\partial x^i}, \quad \text{or}$$
$$v^i = f^i(x,u),$$

the family of control equations. In many cases the fibre bundle $\kappa : \mathbb{U} \to M$ is a trivial bundle, i.e., equals a product $M \times U$ for some input space U. The set of available controls are no longer state dependent. In this case an alternative but equivalent global description of the continuous time nonlinear control system (2.36) is provided by a family of vector fields on the state space manifold M, parameterized by the inputs $u \in U$.

A common point of view in the literature of geometric optimal control theory is to think of a control system as a family of assigned vector fields on a manifold

$$\mathcal{F} = \{F_u(\cdot) = f(\cdot, u)\}_{u \in U}. \tag{2.38}$$

We now need the concept of an admissible control and of trajectory of a control system.

Definition 2.2.1. *The set*

$$\mathcal{U}_{[0,\mathtt{T}]} = \{u : [0, \mathtt{T}] \to \mathbb{U}_{x(\cdot)} \mid u(\cdot) \text{ is piecewise smooth}, \ \ \mathbb{U}_x = \kappa^{-1}(x)\}$$

denotes the collection of all admissible controls.

Definition 2.2.2. *A trajectory of (2.36) corresponding to $u(\cdot)$ is a map $\gamma(\cdot)$: $[0, \mathtt{T}] \to M$, Lipschitz continuous on every chart, such that (2.36) is satisfied for almost every $t \in [0, \mathtt{T}]$.*

Definition 2.2.3. *The set*

$$\mathcal{R}_{x_0}(\mathtt{T}) := \{x \in M \ : \ \exists t \in [0, \mathtt{T}] \text{ and } \gamma : [0, t] \to M \text{ such that } \gamma(0) = x_0, \gamma(t) = x\}$$

is called the reachable set within time $\mathtt{T} > 0$.

We introduce performance measures which are defined as cost functionals on the state-control trajectory, as follows

$$J(x(.), u(.)) := \int_0^{\mathtt{T}} L(x(t), u(t))dt.$$

The optimization of these measures biases the search of feasible trajectories towards cost-efficient trajectories. The above cost functional is called a Lagrange functional. Cost functionals of the type

$$J(x(.), u(.)) := \phi(x(\mathtt{T}), u(\mathtt{T}))$$

where the optimizer receives a payoff at the terminal-time are called Meyer functionals. Allowing both gives Bolza functionals.

Problem 2.2.4. (Optimal Control Problem) *Given the state dynamics $\dot{x} = f(x, u)$, find a control trajectory $u(t),\ t \in [0, \mathrm{T}]$ among all admissible control trajectories which minimizes the functional $J(u)$ and the corresponding solution $x_u(t)$ of the Cauchy problem satisfies the boundary condition $x_u(\mathrm{T}) = x_{\mathrm{T}}$. The problem can be stated mathematically as following:*

$$\dot{x} = f(x, u(t)), \quad x \in M, \quad u(t) \in \mathcal{U}_{[0,\mathrm{T}]} \tag{2.39}$$

$$x(0) = x_0, \tag{2.40}$$

$$x(\mathrm{T}) = x_{\mathrm{T}} \tag{2.41}$$

$$\min_{u(t)} J(u) = \min_{u(t)} \int_0^{\mathrm{T}} L(x(t), u(t))dt \tag{2.42}$$

2.2.2 The Maximum Principle

To formulate the theorem, we shall consider, in addition to the fundamental system another system of equations known as the *conjugate* system in the auxiliary variables $\psi_0, \psi_1, \ldots, \psi_n$:

$$\frac{d\psi^i}{dt} = -\sum_{\alpha=0}^{n} \frac{\partial f^\alpha(x,u)}{\partial u^\alpha}\psi_\alpha, \quad i = 0, 1, \ldots, n.$$

This system has the following geometric interpretation: In the space of vectors $(\psi_0, \psi_1, \ldots, \psi_n)$ dual (conjugate) to the space of vectors $x^0, x^1, \ldots, x^n$, consider the hyperplane

$$\sum_{\alpha=0}^{n} \psi_\alpha x^\alpha = c = \text{ const}$$

passing through the initial point $(0, x_0^1, \ldots, x_0^n)$. Then the conjugate system describes the transport of this hyperplane along the trajectories corresponding to solutions of the state system. In other words, if the ψ_i satisfy the conjugate system and the x^i satisfy the state equations for $t_0 \le t \le t_1$, then

$$\sum_{\alpha=0}^{n} \psi_\alpha x^\alpha = c \quad t_0 \le t \le t_1.$$

If we choose an admissible control $u(t), t_0 \le t \le t_1$, and have the corresponding phase trajectory $x(t)$ of the fundamental system with initial condition $x(t_0) = x_0$, then the adjoint system takes the form

$$\frac{d\psi^i}{dt} = -\sum_{\alpha=0}^{n} \frac{\partial f^\alpha(x(t),u(t))}{\partial u^\alpha}\psi_\alpha, \quad i = 0, 1, \ldots, n.$$

This system is linear and homogeneous. Therefore, for any initial conditions, it admits the unique solution

$$\psi = (\psi_0, \psi_1, \ldots, \psi_n)$$

for the ψ_i. Just as the solution $x(t)$ of the fundamental system, the solution of the adjoint system consists of continuous functions $\psi_i(t)$ which have everywhere, except at a finite number of points, continuous derivatives with respect to t.

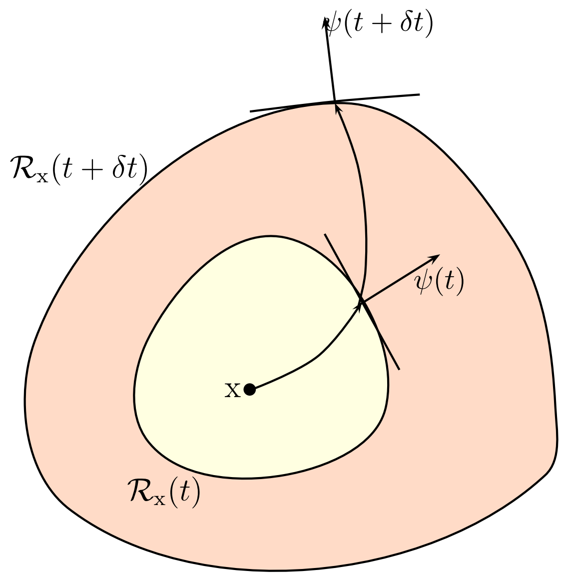

Figure 2.2: Reachable set

We shall now combine the fundamental system and the adjoint system into one entry. To do so, we introduce the Hamiltonian $\mathcal{H}$ of the variable $x^1, \ldots, x^n, \psi_0, \ldots, \psi_n, u^1, \ldots, u^r$:

$$\mathcal{H}(\psi, x, u) = (\psi, f(x, u)) = \sum_{\alpha=0}^{n} \psi_\alpha f^\alpha(x, u).$$

We can rewrite the Hamiltonian in the form of the Hamiltonian system:

$$\frac{dx^i}{dt} = \frac{\partial \mathcal{H}}{\partial \psi_i}, \qquad i = 0, 1, \ldots, n, \tag{2.43}$$

$$\tag{2.44}$$

$$\frac{d\psi_i}{dt} = -\frac{\partial \mathcal{H}}{\partial x^i}, \qquad i = 0, 1, \ldots, n. \tag{2.45}$$

Thus, taking an arbitrary admissible control $u(t), t_0 \leq t \leq t_1$, and the initial condition $x(t_0) = x_0$, we can find the corresponding trajectory $x(t) = (x^0(t), \ldots, x^n(t))$. After that we can find the solutions of the Hamiltonian system

$$\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$$

corresponding to the functions $u(t)$ and $x(t)$.

For fixed values of ψ and x, the function $\mathcal{H}$ becomes a function of the parameter $u \in U$. Let us denote the least upper bound of the values of this function by $\mathcal{M}(\psi, x)$:

$$\mathcal{M}(\psi, x) = \sup_{u \in U} \mathcal{H}(\psi, x, u).$$

If the continuous function $\mathcal{H}$ achieves its upper bound on U, then $\mathcal{M}(\psi, x)$ is the maximum of the values of $\mathcal{H}$, for fixed ψ and x. We now state the necessary condition for optimality, whose principle content is called the *Maximum principle*.

Theorem 2.2.5. *Let $u(t), t_0 \leq t \leq t_1$, be an admissible control such that the corresponding trajectory $x(t)$ which begins at the point x_0 at the time t_0 passes, at some time t_1, through a point on the line Π. In order that $u(t)$ and $x(t)$ be* *optimal it is necessary that there exist a nonzero continuous vector function* $\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$ *corresponding to* $u(t)$ *and* $x(t)$*, such that:*

1. *for every* $t, t_0 \leq t \leq t_1$*, the function* $\mathcal{H}(\psi(t), x(t), u)$ *of the variable* $u \in U$ *attains its maximum at the point* $u = u(t)$*:*

$$\mathcal{H}(\psi(t), x(t), u(t)) = \mathcal{M}(\psi(t), x(t));$$

2. *at the terminal time* t_1 *the relations*

$$\psi_0(t_1) \leq 0, \quad \mathcal{M}(\psi(t_1), x(t_1)) = 0$$

are satisfied. Furthermore, it turns out that if $\psi(t), x(t)$*, and* $u(t)$ *satisfy the Hamiltonian system, and condition (1), the time functions* $\psi_0(t)$ *and* $\mathcal{M}(\psi(t), x(t))$ *are constant. Thus, condition (2) may be verified at any time* $t, t_0 \leq t \leq t_1$*, and not just at* t_1*.*

2.2.3 Maximum principle for minimum-time problem

We shall now derive an analogous necessary condition for the minimum-time problem. To do this, it is only necessary to set $f^0(x, u)$ equal to 1. The hamiltonian $\mathcal{H}$ then takes the form

$$\mathcal{H} = \psi_0 + \sum_{\alpha=1}^{n} \psi_\alpha f^\alpha(x, u).$$

Introducing the n-dimensional vector $\psi = (\psi_1, \ldots, \psi_n)$ and the function

$$H = \sum_{\alpha=1}^{n} \psi_\alpha f^\alpha(x, u),$$

we can write the corresponding Hamiltonian system

$$\frac{dx^i}{dt} = \frac{\partial H}{\partial \psi_i}, \qquad i = 1, \ldots, n, \tag{2.46}$$

$$\tag{2.47}$$

$$\frac{d\psi_i}{dt} = -\frac{\partial H}{\partial x^i}, \qquad i = 1, \ldots, n. \tag{2.48}$$

For fixed values of ψ and x, H is a function of u. We denote the upper bound of the values of this function by $M(\psi, x)$:

$$M(\psi, x) = \sup_{u \in U} H(\psi, x, u).$$

Because $H(\psi, x, u) = \mathcal{H}(\psi, x, u) - \psi_0$, we obtain

$$M(\psi, x) = \mathcal{M}(\psi, x) - \psi_0,$$

and therefore

$$H(\psi(t), x(t), u(t)) = M(\psi, (t), x(t)) = -\psi_0 \geq 0.$$

Thus we obtain the following theorem.

Theorem 2.2.6. *Let $u(t), t_0 \leq t \leq t_1$, be an admissible control which transfers the phase point from x_0 to x_1, and let $x(t)$ be the corresponding trajectory, so that $x(t_0 = x_0)$, $x(t_1) = x_1$. In order that $u(t)$ and $x(t)$ be time-optimal it is necessary that there exist a nonzero, continuous vector function $\psi(\mathbf{t}) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$ corresponding to $u(t)$ and $x(t)$, such that:*

1. *for all* $t, t_0 \leq t \leq t_1$*, the function* $H(\psi(t), x(t), u)$ *of the variable* $u \in U$ *attains its maximum at the point* $u = u(t)$*:*

$$H(\psi(t), x(t), u(t)) = M(\psi(t), x(t));$$

2. *at the terminal time* t_1 *the relation*

$$M(\psi(t_1), x(t_1)) \geq 0$$

are satisfied. Furthermore, it turns out that if $\psi(t), \mathbf{x}(t)$*, and* $u(t)$ *satisfy the Hamiltonian system, and condition (1), the time function* $M(\psi(t), x(t))$ *is constant. Thus, condition (2) may be verified at any time* $t, t_0 \leq t \leq t_1$*, and not just at* t_1*.*

Moreover all optimal control problems could be reduced to minimum-time problems by re-parameterizing time t with arc-length. Furthermore one can derive the necessary conditions for state and control constrained problems. These constraints are given as algebraic equality or inequality constraints:

$$g(x, u, t) = 0, \tag{2.49}$$

$$h(x, u, t) \leq 0, \tag{2.50}$$

which are path constraints, i.e., they hold true in the interval $[t_0, t_f]$. Some of these constraints could be of boundary type, namely

$$\phi(t_f, x(t_f)) \leq 0, \tag{2.51}$$

where t_f is the final time.

2.3 Geodesics

2.3.1 Riemannian Metric

Suppose M is a smooth manifold, and G is a symmetric covariant tensor field of rank 2 on M. If $(U; x^i)$ is a local coordinate system on M, then the tensor field G can be expressed as

$$G = g_{ij}(x)\, dx^i \otimes dx^j, \tag{2.52}$$

on U, where $g_{ij} = g_{ji}$ is a smooth function on U, usually defined to be $\left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle \bigg|_x$. G provides a bilinear function on T_xM at every point $x \in M$. Suppose $X, Y \in \Gamma(TM)$, then

$$\begin{aligned} G(X,Y) &= \langle X, Y \rangle \\ &= \left\langle X^i \frac{\partial}{\partial u^i}\, ,\; Y^j \frac{\partial}{\partial u^j} \right\rangle \\ &= X^i Y^j \left\langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \right\rangle \\ &= g_{ij} X^i Y^j. \end{aligned} \tag{2.53}$$

We say that the tensor G is nondegenerate at the point x if, whenever $v \in T_xM$ and

$$G(v, w) = 0, \qquad \text{for all } w \in T_pM, \tag{2.54}$$

it must be true that $v = 0$. This implies that G is nondegenerate at x iff the system of linear equations

$$g_{ij}(x)\, v^i = 0, \qquad 1 \leq j \leq n \tag{2.55}$$

has zero as its only solution, i.e., $\det g_{ij}(p) \neq 0$. If for all $v \in T_xM$ we have

$$G(v, v) \geq 0, \tag{2.56}$$

and the equality holds only if $v = 0$, then we say G is positive definite at x, which implies that G is necessarily nondegenerate.

For a generalized Riemannian manifold M, G specifies an inner product on the tangent space T_xM at every point $x \in M$. For any $X, Y \in T_xM$, let

$$X \cdot Y = G(X, Y) = g_{ij}\, X^i Y^j. \tag{2.57}$$

When G is positive definite, it is meaningful to define length of a tangent vector and the angle between two tangent vectors at the same point, i.e.,

$$\|X\| = (g_{ij}\, X^i Y^j)^{\frac{1}{2}}, \tag{2.58}$$

$$\cos \angle(X, Y) = \frac{X \cdot Y}{\|X\| \cdot \|Y\|}. \tag{2.59}$$

Thus a Riemannian manifold is a differentiable manifold which has a positive definite inner product on the tangent space at every point. The inner product is required to be smooth, i.e., if X, Y are smooth tangent vector fields, then $X \cdot Y$ is a smooth function on M.

The differential 2-form

$$ds^2 = g_{ij}\, dx^i dx^j \tag{2.60}$$

is independent of the choice of the local coordinate system and is usually called the metric form or Riemannian metric. ds is precisely the length of an infinitesimal tangent vector, and is called the element of arc length.

Suppose $C : x^i = x^i(t), \quad t_0 \leq t \leq t_1$, is a continuous and piecewise smooth parameterized curve on M. Then the arc length of C is defined to be

$$s = \int_{t_0}^{t_1} \left(g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} \right)^{\frac{1}{2}} dt. \tag{2.61}$$

Theorem 2.3.1. *There exists a Riemannian metric on any n-dimensional smooth manifold M.*

Isomorphism

With the help of a fundamental tensor, we may identify a tangent space with a cotangent space, and hence a contravariant vector and a covariant vector can be viewed as different expressions of the same vector. In fact, if $X \in T_xM$, let

$$\alpha_X(Y) = G(X, Y), \qquad Y \in T_xM. \tag{2.62}$$

Then α_X is a linear functional on T_xM, i.e., $\alpha_X \in T_x^*M$.

Conversely, since G is nondegenerate, any element of T_p^*M can be expressed in the form α_X. Thus α establishes an isomorphism between T_xM and T_x^*M. Componentwise, if

$$X = X^i \frac{\partial}{\partial x^i}, \qquad \alpha_X = X_i dx^i, \tag{2.63}$$

then we obtain from (2.62) that

$$X_i = g_{ij} X^j, \qquad X^j = g^{ij} X_i. \tag{2.64}$$

In general, if (t^i_{jk}) is a $(1,2)$-type tensor, then

$$t_{ijk} = g_{il} t^l_{jk}, \qquad t^{ij}_k = g^{jl} t^i_{lk} \tag{2.65}$$

are $(0,3)$-type and $(2,1)$-type tensors, respectively.

2.3.2 Levi-Civita Connection

Metric-compatible connection

Definition 2.3.2. *Suppose (M, G) is an n-dimensional Riemannian manifold, and D is an affine connection on M. If*

$$DG = 0, \tag{2.66}$$

then D is called a metric-compatible *connection on (M, G).*

Condition (2.66) means that the fundamental tensor G is parallel with respect to metric-compatible connections. If the connection matrix of D under the local coordinates x^i is $\omega = (\omega_i^j)$, then

$$DG = (dg_{ij} - \omega_i^k g_{kj} - \omega_j^k g_{ik}) \otimes dx^i \otimes dx^j.$$

Thus (2.66) is equivalent to

$$\begin{aligned} dg_{ij} &= \omega_i^k g_{kj} - \omega_j^k g_{ik}, \\ &= \omega_{ij} + \omega_{ji} \end{aligned} \tag{2.67}$$

The geometric meaning of metric-compatible connections is that parallel translations preserve the metric. In particular, on a Riemannian manifold, the length of a tangent vector and the angle between two tangent vectors are invariant under parallel translations.

In fact, if $X(t), Y(t)$ are parallel vector fields along the curve $C : x^i = x^i(t)$, $(1 \leq i \leq n)$ with respect to a metric-compatible connection, then

$$\frac{dX^i}{dt} + \Gamma^i_{jk} X^j \frac{dx^k}{dt} = 0, \tag{2.68}$$

$$\frac{dY^i}{dt} + \Gamma^i_{jk} Y^j \frac{dx^k}{dt} = 0. \tag{2.69}$$

Hence

$$\begin{aligned} \frac{d}{dt}(g_{ij} X^i Y^j) &= \frac{dg_{ij}}{dt} X^i Y^j + g_{ij} \left(\frac{dX^i}{dt} Y^j + X^i \frac{dY^j}{dt} \right), \\ &= \left(\frac{dg_{ij}}{dt} - g_{ik} \Gamma^k_{jh} \frac{du^h}{dt} - g_{jk} \Gamma^k_{ih} \frac{du^h}{dt} \right) X^i Y^j, \quad \text{by (2.68), (2.69)}, \\ &= 0, \qquad \text{using (2.67)}. \end{aligned}$$

Therefore, along C, we have

$$g_{ij}X^iX^j = \text{ const } . \tag{2.70}$$

Torsion-free connection

The torsion tensor T of the affine connection D is $(1,2)$-type tensor. Being a $(1,2)$-type tensor, T can be viewed as a map

$$T : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM).$$

Suppose X, Y are any two tangent vector fields on M, then $T(X,Y)$ is a tangent vector field on M with local expression

$$T(X,Y) = T^k_{ij}X^iY^j\frac{\partial}{\partial u^k}.$$

From which it follows that

$$T(X,Y) = D_XY - D_YX - [X,Y]\,. \tag{2.71}$$

Definition 2.3.3. *If the torsion tensor of an affine connection D is zero,i.e.,*

$$D_XY - D_YX = [X,Y]\,, \tag{2.72}$$

the connection is said to be torsion-free *(or* symmetric*).*

Theorem 2.3.4. *(Fundamental Theorem of Riemannian Geometry). Suppose M is an n-dimensional generalized Riemannian manifold. Then there exists a unique torsion-free and metric compatible connection on M, called the Levi-Civita connection of M, or the Riemannian connection of M.*

Proof. Suppose D is a torsion-free and metric compatible connection on M. Denote the connection matrix of D under the local coordinates u^i by $\omega = (\omega_i^j)$, where

$$\omega_i^j = \Gamma_{ik}^j du^k.$$

Then we have

$$\begin{aligned} dg_{ij} &= \omega_i^k g_{kj} + \omega_j^k g_{ki}, \\ \Gamma_{ik}^j &= \Gamma_{ki}^j. \end{aligned}$$

Denote

$$\Gamma_{ijk} = g_{lj}\Gamma_{ik}^l, \qquad \omega_{ik} = g_{lk}\omega_i^l.$$

Then it follows that

$$\begin{aligned} \frac{\partial g_{ij}}{\partial u^k} &= \Gamma_{ijk} + \Gamma_{jik}, \\ \Gamma_{ijk} &= \Gamma_{kji}. \end{aligned}$$

Cycling the indices, we get

$$\begin{aligned} \frac{\partial g_{ik}}{\partial u^j} &= \Gamma_{ikj} + \Gamma_{kij}, \\ \frac{\partial g_{jk}}{\partial u^i} &= \Gamma_{jki} + \Gamma_{kji}. \end{aligned}$$

We then obtain

$$\begin{aligned} \Gamma_{ikj} &= \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial u^j} + \frac{\partial g_{jk}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k}\right) \\ \Gamma_{ij}^k &= \frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial u^j} + \frac{\partial g_{jl}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^l}\right). \end{aligned}$$

Thus we see that the torsion-free and metric-compatible connection is determined uniquely by the metric tensor.

Conversely, the Γ_{ij}^k defined above indeed satisfy the transformation equation for connection coefficients under a change of local coordinates. Hence they define an affine connection D on M. Thus D is a torsion-free and metric-compatible connection on M. □

Riemannian connection in Arbitrary Frame Fields

Here we redo the Riemannian connections in arbitrary frame field instead of the natural frame field. Suppose $(e_1, \ldots, e_n)$ is a local frame field with coframe field $(\theta^1, \ldots, \theta^n)$. Let

$$De_i = \theta_i^j e_j,$$

where $\theta = (\theta_i^j)$ is the connection matrix of the connection D with respect to the frame field $(e_1, \ldots, e_n)$. Hence by structure equations of connections we know that the statement that:

1. D is a torsion-free connection is equivalent to the statement that the θ_j^i satisfy

the equations

$$d\theta^i - \theta^j \wedge \theta^i_j = 0.$$

2. D is a metric-compatible connection is given by

$$dg_{ij} = \theta^k_i g_{kj} + \theta^k_j g_{ik},$$

where $g_{ij} = G(e_i, e_j)$, and the metric form is $ds^2 = g_{ij}\theta^i\theta^j$.

Theorem 2.3.5. *Suppose (M, G) is a generalized Riemannian manifold, and $\{\theta^i,\ 1 \leq i \leq n\}$ is a set of differential 1-forms on a neighborhood $U \subset M$ which is linearly independent everywhere. Then there exists a unique set of n^2 differential 1-forms θ^k_j on U such that*

$$d\theta^i = \theta^j \wedge \theta^i_j,$$

$$dg_{ij} = \theta^k_i g_{kj} + \theta^k_j g_{ik},$$

where the g_{ij} are the components of G with respect to the local coframe field $\{\theta^i\}$, i.e.,

$$G = g_{ij}\theta^i \otimes \theta^j.$$

2.4 Optimal Control on Manifolds

So far, the state spaces have been open subsets of Euclidean spaces, and the formulations of the Maximum Principle have depended on this fact. We now survey some of the literature that lift the Maximum Principle on to manifolds, intrinsically. Sussmann [67], [68], mainly describes the Hamiltonian lift which is symplectic in nature and connections along curves expressed in terms of Lie brackets, Lopez [49], introduces local Finsler metrics on the tangent space in the presence of control constraints, and Udriste [71], improves upon [49], by applying Okubo's method for deriving metrics. We now give a brief outline of each of these techniques.

2.4.1 Hamiltonian Lift

One way to prove invariance of the adjoint equation is to use the canonical symplectic structure of the of the cotangent bundle T^*M of a manifold M. This structure gives rise to a canonical fiber-preserving isomorphism $J : T^*(T^*M) \to T(T^*M)$, which enables us to assign to every covector $\omega \in T^*_z(T^*M)$ a tangent vector $J_z(\omega) \in T_z(T^*M)$. In particular, if $H : T^*M \to \mathbb{R}$ is a function, and $z \in T^*M$ is a point where H is differentiable, then there is a well defined tangent vector $\vec{H}(z) \in T_zT^*M$, given by $\vec{H}(z) = J_z(dH(z))$.

Now suppose that we are given a control system $\dot{x} = f(x, u, t)$ on a manifold M, and a reference trajectory-control pair $\gamma_* = (\xi_*, \eta_*)$ such that f_{η_*} satisfies a C^1-Caratheodary condition near ξ_*. For each u, t, define

$$\mathcal{H}_{u,t}(x, p) := H(x, u, p, t) = p \cdot f(x, u, t). \tag{2.73}$$

Then the function $\mathcal{H}_{\eta^*(t),t}$ is of class C^1 for almost every t. The system

$$\dot{\xi}(t) = f(\xi(t), \eta^*(t), t), \tag{2.74}$$

$$-\dot{\psi}(t) = \psi(t)\frac{\partial f}{\partial x}(\xi(t), \eta^*(t), t) \tag{2.75}$$

is then equivalent to

$$\dot{\zeta}(t) = \vec{\mathcal{H}}_{\eta^*(t),t}(\zeta(t)), \tag{2.76}$$

where $\zeta = (\xi(t), \psi(t))$.

Suppose if $\Sigma = (M, U, I, \mathcal{U}, f)$ is a control system on a manifold M, such that for every $\eta \in \mathcal{U}$ the corresponding vector field f_η satisfies a C^1-Caratheodary condition, then we can define

$$f^*(z, u, t) \;=\; \mathcal{H}_{u,t}(z), \qquad \text{for} \quad (z, u, t) \in T^*M \times U \times I, \tag{2.77}$$

and obtain a new control system $\Sigma^* = (T^*M, U, I, \mathcal{U}, f^*)$ with state space T^*M, called the Hamiltonian lift of Σ. Then the necessary condition given by the Maximum Principle can be viewed as stating that the reference trajectory-control pair $\gamma^* = (\xi^*, \eta^*)$ must be the projection of a reference trajectory-control pair $\Gamma^* = (\gamma^*, \eta^*)$ of the Hamiltonian lift such that Γ^* has some special properties, namely, non-triviality, Hamiltonian maximization, and the transversality condition.

2.4.2 Connections along curves

We now define a connection along a curve ξ, such that $\dot{\xi}(t) = f(\xi(t), t)$ for almost every t. If ξ is an integral curve of a time-varying vector field $f(x,t)$, then the pair (ξ, f) gives rise in a canonical way to a connection $\nabla_{\xi,f}$ along ξ (such that if X is a smooth vector field on M, and $X \circ \xi$ is the vector field along ξ defined by $(X \circ \xi)(t) = X(\xi(t))$) given by

$$\nabla_{\xi,F}(X \circ \xi)(t) = [f_t, X](\xi(t)), \qquad \text{for a.e. } t \in \mathcal{D}(\xi). \tag{2.78}$$

Relative to a local coordinate chart , we then have

$$(\nabla_{\xi,f}(X))^j = \frac{d}{dt}(X^j \circ \xi)(t) - \sum_{i=1}^{n} \left(\frac{\partial f_t^j}{\partial x^i} X^i \right)(\xi(t)), \quad \text{for a.e. } t, \tag{2.79}$$

$$= \frac{\partial X^j}{\partial x^i}\dot{\xi}(t) - \sum_{i=1}^{n} \left(\frac{\partial f_t^j}{\partial x^i} X^i \right)(\xi(t)), \tag{2.80}$$

where $\dot{\xi}(t) = f(\xi(t), t)$.

A connection along a curve induces an operator of "parallel translation" along the curve. Precisely, if $\xi : [a,b] \to M$ is a curve and ∇ is a connection along ξ, then a vector field θ along ξ is ∇-parallel if

$$\nabla\theta \equiv 0. \tag{2.81}$$

If $t \in [a,b]$ and $v \in T_{\xi(t)}M$, then there is a unique vector field θ along ξ which is ∇-parallel and satisfies $\theta(t) = v$.

Furthermore, a connection ∇ along a curve, which is defined as a differential operator on vector fields, induces in a standard way differential operators on fields of covectors and on fields of higher-order tensors. Similarly, if $\zeta \in \Gamma(T^*M, \xi)$, and X is a vector field near $\xi(t)$, we have

$$(\nabla_{\xi,f}\zeta).X(p) = \frac{d}{ds}\bigg|_{s=t} \zeta(s).X(\xi(s)) - \zeta(t).(\nabla_{\xi,f}X) \tag{2.82}$$

$$= \frac{d}{ds}\bigg|_{s=t} \zeta(s).X(\xi(s)) - \zeta(t).[f_t, X](\xi(t)). \tag{2.83}$$

If θ is a field of vectors and ω a field of covectors along ξ, then the product rule

$$\frac{d}{dt}(\omega(t) \cdot \theta(t)) \ = \ \nabla\omega(t) \cdot \theta(t) + \omega(t) \cdot \nabla\theta(t), \tag{2.84}$$

holds. In particular, if θ and ω are parallel transported, then the product $\omega(t)\cdot\theta(t)$is a constant along ξ.

If ξ is an integral curve of a time-varying vector field f, and $\nabla = \nabla_{\xi,f}$, then the equations $\nabla\theta = 0$ and $\nabla\omega = 0$ are the variational and adjoint equations along (ξ, f), respectively. In the setting of the maximum principle, when f is the vector field f_{η^*} corresponding to the reference control η^*, the adjoint equation that appears in the maximum principle is precisely the equation

$$\nabla_{\xi^*, f_{\eta^*}}\psi \equiv 0. \tag{2.85}$$

2.4.3 Metrics on Tangent Spaces

In [49] Finsler metrics on the tangent space of optimal control problems were constructed. Let M be an n-dimensional C^∞ manifold(state space), $(\mathbb{U}, \kappa, M)$ be a control fibre bundle, and (TM, τ, M) be the tangent bundle. Let $x^i,\ i = 1, \ldots, n$ be the components of the point $x \in M$, called state variables. Then $(x^i, u^a),\ i = 1, \ldots, n;\ a = 1, \ldots, k$ are adapted coordinates in $\mathbb{U}$, where u^α are the controls, and (x^i, v^i) are natural coordinates in TM.

Let $X : \mathbb{U} \to TM,\ X = X^i(x, u)\frac{\partial}{\partial x^i}$ be a C^∞ fibred mapping, over the identity in the state manifold M, which produces a continuous control system

$$\frac{dx^i}{dt} = X^i(x, u), \quad i = 1, \ldots, n, \tag{2.86}$$

The evolution of the state manifold M is totally characterized by the image set $S = Im(X) \subset TM$ which is described by the control equations (refeq:lopez1). Pointwise we have the set $S_x = \{v \mid \exists u \in \mathbb{U}_x = \kappa^{-1}(x), v = f(x, u)\}$.

The chief objects involved in the metric construction are the set of allowed directions, a cone bundle $V \subset TM$, and a Finslerian (homogeneous) function defined on F. These objects are defined at each point of the state space M. Once the state x is fixed, the maximality condition of the Maximum Principle determines a subset $\mathbb{U}_x^*$ of $\mathbb{U}_x = \kappa^{-1}(x)$ of possible optimal controls

$$\mathbb{U}_x^* = \{(x, u_0) \in \mathbb{U}_x \mid \exists (x, p) \in T_x^*M \quad \text{such that} \quad H(x, p, u_0) \geq H(x, p, u),\ \ \forall u \in \mathbb{U}_x\}\,, \tag{2.87}$$

from which the optimal control $U^*(x, p)$ is determined for every p.

The image subset $S_x^* = X(\mathbb{U}_x^*) \subset S_x$ contains some of the longest allowed velocity vectors, because for a given momentum covector $p_i dx^i$, the Hamiltonian is maximal for the vector $f^i(x,u)\frac{\partial}{\partial x^i}$ with the greatest projection, $\max\langle p, f(x,u)\rangle$, i.e., the longest in some particular ray direction. Upon reparametrization one need to only look at unit vectors in each direction, the set of all unit vectors form the indicatrix. So we define S_x^* as follows

$$S_x^* = X(\mathbb{U}_x^*) = \{(x,v) \in T_xM \mid v \in S_x \quad \text{and} \quad \lambda v \notin S_x, \;\; \forall \lambda > 1\}\,. \tag{2.88}$$

The invariance under reparametrization needed to define a metric can be obtained by considering the cone V_x of rays generated by elements of S_x

$$V_x = \Big\{(x,v) \mid \exists (x,v_0) \in S_x^0 \quad \text{such that} \quad v = \lambda v_0,\; \lambda > 0\Big\}\,. \tag{2.89}$$

The norm of the velocity is now defined to obtain a length equal to the original cost; we associate to every $(x,v) \in V_x$ the norm λ, which is the factor between it and the element of S_x^0 in the same ray

$$F_x : V_x \longrightarrow \mathbb{R}^+, \quad F_x(v) = \lambda, \text{ where } \; v = \lambda v_0, \text{ with } \; v_0 \in S_x^0. \tag{2.90}$$

An optimal curve $\rho(t) = (x(t), u^*(t))$, with cost $\int_0^T dt = T$, will have, once arbitrarily reparametrized by $t(\tau)$ ($\frac{dt}{d\tau} > 0$), the length is given

$$\int F(x(\tau), v(\tau))d\tau = \int \lambda(\tau)d\tau = \int \frac{dt}{d\tau}d\tau = \int dt = T, \tag{2.91}$$

because

$$v(\tau) = \frac{dx}{d\tau} = \frac{dx}{dt}\frac{dt}{d\tau} = v_0\frac{dt}{d\tau}, \text{ so that } F(x,v) = \lambda = \frac{dt}{d\tau}.$$

The set $S_x \subset V_x$ is the indicatrix of the defined metric.

Example 4. Here we describe construction of the metric for the following two-dimensional time optimal control problem:

$$\frac{dx}{dt} = 2 + u^1 \cos(u^2), \quad \frac{dy}{dt} = u^1 \sin(u^2), \quad 0 \le u^1 \le 1, \ \ u^2 \in S^1. \tag{2.92}$$

It is quite obvious that for any given point $(x_0, y_0) \in \mathbb{R}^2$, the subset $S_{(x_0,y_0)}$ is the unit disk centered at $(2, 0)$. The subset of possible optimal allowed velocities $S^*_{(x_0,y_0)}$ is determined by $u^1 = 1$ and $-\frac{2\pi}{3} \le u^2 \le \frac{2\pi}{3}$, the boundary of the disk made of the longest vectors. $D_{(x_0,y_0)}$ is determined by the conditions $|\frac{v_y}{v_x}| \le \frac{1}{\sqrt{3}}$ and $v_x > 0$. The Finslerian function, the factor between elements of $D_{(x_0,y_0)}$ and the elements of $S^0_{(x_0,y_0)}$ on the same ray, is explicitly given by

$$F_{(x_0,y_0)}(v_x, v_y) = \frac{v_x^2 + v_y^2}{2v_x + \sqrt{(v_x^2 - 3v_y^2)}}, \tag{2.93}$$

Okubo's Technique

Let $f : TM \to R$ be a C^∞ function on $TM \setminus \{0\}$, whose restriction $v \to f(x, v)$ has no critical point. Then

$$I_x : f(x, v) = 0 \tag{2.94}$$

is a hypersurface of T_xM, for each $x \in M$. On the other hand, there exists an implicit function $F : TM \setminus \{0\} \to R$ which is C^∞ and positive homogeneous of degree one with respect to v such that $I_x : F(x, v) = 1$. In this sense I_x is an indicatrix. Okubo [51], has shown that the implicit function F is a solution of the equation

$$f\left(x, \frac{v}{F(x, v)}\right) = 0. \tag{2.95}$$

Conversely, given the Finsler function F there exists at least two different functions providing the same indicatrix.

Udriste [71], improves upon [49] by devising the Finsler metrics on local tangent spaces in the following way: let V_x be a cone in T_xM, i.e., V_x is a subset of T_xM which is invariant with respect to positive homotheties. If

$$f(x,v) = \begin{cases} f_1(x,v), & v \in V_x \subset T_xM \\ f_2(x,v), & v \in T_xM \setminus V_x \end{cases} \tag{2.96}$$

has suitable properties, then one can produce a Finsler function $F = \sqrt{(}F_1^2 + F_2^2)$, which satisfies

$$f\left(x, \frac{v}{F(x,v)}\right) = 0, \qquad \forall v \in T_xM \tag{2.97}$$

via

$$f_1\left(x, \frac{v}{F_1(x,v)}\right) = 0, \quad \forall v \in V_x \tag{2.98}$$

$$f_2\left(x, \frac{v}{F_2(x,v)}\right) = 0, \quad \forall v \in T_xM \setminus V_x. \tag{2.99}$$

With respect to a time-optimal control problem, the cone V_x is as given in (2.89), then the function F_1 is defined as follows

$$F_{1x} : V_x \to R, \;\; F_{1x}(v) = \lambda, \;\; \text{where} \;\; v = \lambda v_0, \; v_0 \in S_{1x} \tag{2.100}$$

where λ is the factor between v with $v \in V_x$ and v_0 with $v_0 \in S_{1x}$ in the same ray. If there exists a suitable function $f_{1x} : V_x \to R$ with $S_{1x} = f_{1x}^{-1}(0)$, then F_1 is a solution of (2.98). A second arbitrary indicatrix $S_{2x} : f_2(x,v) = 0$ in $T_xM \setminus V_x$ and a positive homogeneous function of degree one $F_2(x,v)$ given by equation (2.99).

Example 5. Continuing Example 4, the image set S is given by

$$S : (v_x - 2)^2 + v_y^2 \leq 1; \tag{2.101}$$

and the indicatrix S_1 by

$$(v_x - 2)^2 + v_y^2 = 1, \quad v_y \geq \frac{3}{2}, \tag{2.102}$$

which can be written in controls as $u^1 = 1, u^2 \in \left[-\frac{2\pi}{3}, \frac{2\pi}{3}\right]$. The cone V_x is given by

$$V_x : \left|\frac{v_y}{v_x}\right| \leq \frac{1}{\sqrt{3}}, v_x > 0. \tag{2.103}$$

The positive homogeneous solution of

$$\left(\frac{v_x}{F_1} - 2\right)^2 + \left(\frac{v_y}{F_1}\right)^2 = 1 \tag{2.104}$$

is

$$F_1(v_x, v_y) = \frac{v_x^2 + v_y^2}{2v_x + \sqrt{(v_x^2 - 3v_y^2)}}. \tag{2.105}$$

On $T_xM \setminus V_x$, one can build an arbitrary positive homogeneous function F_2 of degree one, and finally $F = \sqrt{(F_1^2 + F_2^2)}$ is a homogeneous function of degree one on T_xM.

Chapter 3

Geometric Optimal Control Theory: Finsler Approach

3.1 Introduction

Riemannian geometry is based on a family of positive definite quadratic forms, one defined at each point x on a manifold M. The family is called the Riemannian metric ground form

$$ds^2 = g_{ij}(x)dx^i \otimes dx^j, \tag{3.1}$$

where the x^i are the local coordinates and $g_{ij}(x)$ are smooth functions of position x. This encapsulates the idea of a locally Euclidean metric for measuring distance, one which varies smoothly with the point x.

In contrast Finsler geometry is based on a family of positive definite quadratic forms, one defined at each point $(x, \dot{x})$ of the tangent bundle of a manifold M. The family is called the Finsler metric ground form

$$ds^2 = g_{ij}(x, \dot{x}) dx^i \otimes dx^j, \tag{3.2}$$

where the form and distance vary smoothly as $(x, \dot{x})$ varies over the tangent bundle of M.

A Riemannian metric is locally Euclidean on M, a Finsler manifold is locally Euclidean on the tangent manifold of M. Infinitisimal Finsler length ds varies smoothly with position x and direction $\dot{x}$, while infinitisimal Riemannian length ds varies smoothly based on position x alone.

Let M be an n-dimensional differentiable manifold with coordinate functions $\{x^j : j = 1, \ldots, n\}$ on a coordinate neighborhood U. The tangent vector of a smooth arc $\gamma : [t_0, t_1] \to U$, referred to a parameter t, has components $v^j = dx^j/dt$. Given a smooth nonnegative function F on the tangent bundle TM of M, an "arc-length" may be assigned to the curve γ, namely

$$s = \int_{t_0}^{t_1} F(x_1, \ldots, x_n; v_1, \ldots, v_n) dt. \tag{3.3}$$

In the language of optimization we refer to F as a cost function and s as the total cost of traversing the curve γ starting at time t_0, ending at time t_1. We assume that the function F is homogeneous of the first degree in the directional arguments $\{v^j\}$, that is for positive λ,. $F(x_1, \ldots, x_n; \lambda v_1, \ldots, \lambda v_n) = \lambda F(x_1, \ldots, x_n; v_1, \ldots, v_n)$. Under these circumstances F is called a Finsler metric function.

The first example of a non-Riemannian Finsler geometry is due to Riemann in his Inaugural lecture of 1851 as he defined Riemannian Geometry on manifolds for the first time. His example was based on quartic forms and does not appear there as connected with the calculus of variations. C. Caratheodory [24], [25] used the Finsler indicatrix, the hypersurface

$$F(x^1, x^2 \dots, x^n; v^1, \dots, v^n) = 1$$

in the tangent space T_pM of a point $p \in M$ with coordinates $\{x^i\}$ in the investigation of discontinuous solutions to variational problems. The classical excess condition of Weierstrass for the integral (3.3) implies the convexity of the indicatrix. The indicatrix generalizes the unit sphere in Riemannian geometry. The metric is essentially determined by the indicatrix. The relation between convex bodies centered at zero and metrics is basic to Minkowski's Geometry of Numbers. The first systematic investigation of the local geometry of manifolds whose tangent spaces are endowed with such indicatrices was carried out in 1918 in the dissertation of P. Finsler [31] under Caratheodory. Concepts such as angle, curvatures of curves and of surfaces were defined by Finsler in terms of a general metric function.

Optimal control in a coordinate free manner on manifolds has been carried out by several authors, notably Sussmann [67], [68], who stated the Pontryagin Maximum Principle in a coordinate independent notation. There the variational and the adjoint equations of the maximum principle are exactly the parallel transport equations of vectors and covectors, respectively, of local connections defined as Lie brackets. Another method mentioned in that literature is the iterated Hamiltonian lift, which has a symplectic flavor. Finsler methods for formulating optimal control have been suggested by several people, including [49], [71]. Lopez [49], suggests that indicatrices are the primary objects of study, possibly because bang-bang controls appear as holes on the indicatrices. But this does not yield computation techniques. Udriste [71] proposed using the Okubo technique(see [4], [8], [28]) for generating metrics. None of these investigations has yet led to algorithms for solving for optimal control, but then coordinate free methods are more for beauty, understanding, generality than for computation.

In 1992-4 Kohn and Nerode observed that every optimal control problem gives rise to a Finsler metric ground form on a suitable manifold. They began the development of optimal control theory based on Finsler Geometry in [43], [44], [45] with the goal in mind of computing near optimal controls in real time. Optimal trajectories are Finsler geodesics, optimal controls are Finsler connections. The principal problem is to compute them, preferably by real time algorithms.

This thesis develops further symbolic and numerical computation tools for determining connections, curvature, torsion, volumes for the eventual purpose of computing close to optimal control for real systems. It incidentally establishes the exact relations between the Finsler geodesic formulation and the Pontryagin maximum principle.

In section 2 we summarize Pontryagin's necessary conditions and describe the Problem of Lagrange and the usual transformation of optimal control problems into the calculus of variations In section 3 we derive the fundamental properties of Finsler spaces along with Finsler Geometry. In section 5 the necessary conditions for the transformed optimal control problem are derived as Euler-Lagrange equations for these metrics. Further theorems for efficient computation of these conditions in symbolic form are proved. In the last section we explicitly derive the geodesic equations in the presence of state equality constraints.

3.2 Lagrange and Pontryagin

3.2.1 Maximum Principle

In optimal control theory we are given a system whose dynamics follow the following under-determined system of ordinary differential equations:

$$\dot{x} = f(x, u), \quad \text{for } x \in M \subseteq \mathbb{R}^n, \; u \in U \subseteq \mathbb{R}^m, \tag{3.4}$$

where $x = (x^i)$ represents the state space variables, $u = (u^i)$ represents the control variables and $f : M \times U \to TM$ is a function. We assume that, for each $x_0 \in \mathbb{R}^n$, the map $u \mapsto f(x_0, u)$ is an embedding of $\mathbb{R}^m$ into $\mathbb{R}^n$, i.e., the controls should appear non-degenerately in the equations.

From the geometric point of view, it can be understood as a fibred mapping

$$X : \mathbb{U} \to M \tag{3.5}$$

from a control fibre bundle $(\mathbb{U}, \kappa, M)$ over the state manifold M to the tangent bundle (TM, π, M). Here the fibres of the control fibre bundle $(\mathbb{U}, \kappa, M)$, $\kappa^{-1}(x)$, $x \in M$ denote the state dependent input spaces. The above geometric description can be best described by the following commutative diagram:

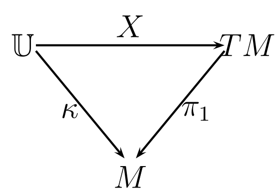

Figure 3.1: Control System: Fibre Bundles

In local coordinates (x^i) in M, adapted coordinates (x^i, u^j) in $\mathbb{U}$, and natural coordinates (x^i, v^i) in TM, the coordinate expression for X is

$$\begin{aligned} X(x,u) &= f^i(x,u)\frac{\partial}{\partial x^i}, \qquad \text{or} \\ v^i &= f^i(x,u), \end{aligned}$$

the family of control equations. In many cases the fibre bundle $\kappa : \mathbb{U} \to M$ is a trivial bundle, i.e., equals a product $M \times U$ for some input space U. The set of available controls are no longer state dependent. In this case an alternative but equivalent global description of the continuous time nonlinear control system (3.4) is provided by a family of vector fields on the state space manifold M, parameterized by the inputs $u \in U$. The optimal control problem could be stated as follows:

Problem 3.2.1. (Optimal Control Problem) *Given the state dynamics* $\dot{x} = f(x,u)$*, find the control trajectory* $u(t),\ t \in [0, \mathtt{T}]$ *among all admissible control trajectories which minimizes the functional* $J(u)$ *and the corresponding solution* $x_u(t)$ *of the Cauchy problem satisfies the boundary condition* $x_u(\mathtt{T}) = x_{\mathtt{T}}$*. The problem can be stated mathematically as following:*

$$\begin{aligned} \dot{x} &= f(x, u(t)), \quad x \in M, \quad u(t) \in \mathcal{U}_{[0,\mathtt{T}]} \\ x(0) &= x_0, \\ x(\mathtt{T}) &= x_{\mathtt{T}} \\ \min_{u(t)} J(u) &= \min_{u(t)} \int_0^{\mathtt{T}} L(x(t), u(t))dt \end{aligned}$$

Pontryagin's maximum principle [58] gives a set of necessary conditions for a state-control trajectory pair $(x(t), u(t))$ to be optimal; by introducing a Hamiltonian function

$$H(x, \lambda, u) = L(x, u) + \lambda_i f^i(x, u), \tag{3.6}$$

where the variables λ_i are adjoint variables. The optimal curves $(x(t), u^*(t))$ must satisfy the control system equations

$$\frac{dx^i}{dt} = \frac{\partial H}{\partial \lambda^i} = f^i(x(t), u^*(t)), \tag{3.7}$$

and there must exist a solution curve for the adjoint differential equations

$$\frac{d\lambda_i}{dt} = -\frac{\partial H}{\partial x^i} = -\frac{\partial L}{\partial x^i}\bigg|_{(x(t),u^*(t))} - \lambda_j \frac{\partial f^j}{\partial x^i}\bigg|_{x(t),u^*(t))} \tag{3.8}$$

with the optimal control u^* satisfying the algebraic condition of maximality

$$H(x, \lambda, u^*) \geq H(x, p, u), \qquad \forall u \in \mathbb{U}_x = \kappa^{-1}(x). \tag{3.9}$$

Equations (3.7), (3.8), and (3.9) together form the maximum principle. For very simple problems one could solve the maximum condition (3.9) to get the controls $u \in U$ as a function of state and adjoint variables. This is a nontrivial operation and usually one resorts to numerical computations.

Our method of introducing a Finsler metric circumvents part of this problem by working on a higher dimensional space whose tangent space includes the tangent space of the original state space, the controls, and the lagrange multipliers. In this way we interpret the controls and adjoint variables as in the tangent space of this higher dimensional space. Once the metric is formed one can derive the Finsler geodesic equation which then become the Euler-Lagrange equations in the Finsler space. It is important to note that the Pontryagin's Maximum Principle are first order necessary conditions in the primal-dual space and hence are neither Euler-Lagrange equations nor the Hamiltonian- Jacobi equations. However, in the augmented space that we propose here (we think for the first time) we can interpret them to be Euler-Lagrange equations. Then we show that these Euler-Lagrange equations are exactly the equations of the Pontryagin's Maximum Principle.

3.2.2 The Problem of Lagrange

In this section our aim is to formulate the optimal control problem as a standard Calculus of Variations problem. This reduction to a CoV problem is to facilitate the introduction of a metric for the optimal control problem which otherwise wouldn't be possible because of the maximization condition in the Pontryagin's Maximum Principle. However, it should be noted that the PMP was derived for admissible controls being measurable and for arbitrary choice of the space of controls U which is why PMP deviates from the ordinary CoV problems, as traditionally, the admissible curves in CoV problems were *piecewise smooth*. Although L.C.Young has devised the notion of *generalized curves* with respect to CoV problems, we pursue in this thesis only the case of piecewise smooth curves. There seems to be little difficulty in extending to the more general case, but we do not do this here.

Following Caratheodory's work [23] on the CoV problems with side-conditions known as the *Problem of Lagrange* (PoL), we now augment the state space in the following manner. Given an autonomous control system $\dot{x} = f(x, u)$, where the states $x = (x^1, \ldots, x^n) \in M \subset \mathbb{R}^n$ and controls $u \in U \subset \mathbb{R}^m$, we introduce an augmented state space $\mathtt{x} = (\mathtt{x}^1, \ldots, \mathtt{x}^{2n+m}) \in \mathbb{R}^{2n+m}$ so that the states x, controls u, and the costate variables $\lambda = (\lambda_1, \ldots, \lambda_n) \in T_x^*M$ of the PMP are redefined in the following way

$$
\begin{aligned}
x^i &= \mathtt{x}^i, && i = 1, \ldots, n, \\
u^i &= \dot{\mathtt{x}}^{n+i}, && i = 1, \ldots, m, \\
\lambda_i &= \dot{\mathtt{x}}^{n+m+i}, && i = 1, \ldots, n,
\end{aligned}
$$

where the controls $u \in U$ and the costate variables $\lambda \in T_x^*M$ are now the time derivatives of the augmented state space $\mathtt{x}$ while the state variables are simply the new states.

We thus have our new state space $\mathrm{x} = (\mathrm{x}^i, \mathrm{x}^{n+j}, \mathrm{x}^{n+m+k})$, where $i, k = 1, \ldots, n$, and $j = 1, \ldots, m$, of dimension $2n + m$. We now formulate the problem of Lagrange as following:

Problem 3.2.2. (Problem of Lagrange). *Given the following Lagrangian*

$$\mathbb{L}(\mathtt{x}, \dot{\mathtt{x}}) = L(\mathtt{x}, \dot{\mathtt{x}}) + \sum_{i=1}^{n} \dot{\mathtt{x}}^{n+m+i}(\dot{\mathtt{x}}^i - f^i(\mathtt{x}, \dot{\mathtt{x}})) \tag{3.10}$$

find a curve $t \mapsto \mathtt{x}^*(t)$ *that minimizes (or maximizes) the following integral*

$$\mathcal{I}(\mathtt{x}(t)) = \int_0^{\mathtt{T}} \mathbb{L}(\mathtt{x}, \dot{\mathtt{x}}) dt \tag{3.11}$$

among all curves $\mathtt{x} : [0, \mathtt{T}] \longrightarrow \mathbb{R}^{2n+m}$ *such that* $\mathtt{x}(0) = \mathtt{x}_0$ *and* $\mathtt{x}(\mathtt{T}) = \mathtt{x}_{\mathtt{T}}$.

We now find the Euler-Lagrange solution of the above CoV problem, i.e., the extrema $\mathtt{x}^*(t)$ solve the equation

$$\frac{d}{dt}\frac{\partial \mathbb{L}}{\partial \dot{\mathtt{x}}^i}(\mathtt{x}, \dot{\mathtt{x}}) = \frac{\partial \mathbb{L}}{\partial \mathtt{x}^i}(\mathtt{x}, \dot{\mathtt{x}}), \qquad i = 1, \ldots, 2n + m. \tag{3.12}$$

Theorem 3.2.3. *The necessary conditions of the Problem of Lagrange are equivalent to that of the Pontryagin Maximum Principle of the regular optimal control problem, i.e.,*

$$\text{PMP} \equiv \text{PoL}$$

Proof. Evaluating the above Euler-Lagrange equation (3.12) for each state variable we get the following:

$$\frac{d}{dt}\begin{bmatrix} \dot{\mathrm{x}}^{n+m+i} \\ \\ 0 \\ \\ \dot{\mathrm{x}}^i - \mathbf{f}^i(\mathrm{x}, \dot{\mathrm{x}}) \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial \mathrm{x}^i}(\mathrm{x}, \dot{\mathrm{x}}) - \dot{\mathrm{x}}^j \frac{\partial f^j}{\partial \mathrm{x}^i}(\mathrm{x}, \dot{\mathrm{x}}) \\ \\ \frac{\partial L}{\partial \dot{\mathrm{x}}^i}(\mathrm{x}, \dot{\mathrm{x}}) - \dot{\mathrm{x}}^j \frac{\partial f^j}{\partial \dot{\mathrm{x}}^i}(\mathrm{x}, \dot{\mathrm{x}}) \\ \\ 0 \end{bmatrix} \tag{3.13}$$

Plugging back the original variables u^i, λ^i for $\dot{\mathrm{x}}^{n+i}$, $\dot{\mathrm{x}}^{n+m+i}$, respectively, we get the following set of equations:

$$\begin{cases} \frac{d\lambda^i}{dt} = \frac{\partial L}{\partial x^i}(x, u) - \lambda^j \frac{\partial f^j}{\partial x^i}(x, u), & i = 1, \ldots, n \\ \\ 0 = \frac{\partial L}{\partial u^i}(x, u) - \lambda^j \frac{\partial f^j}{\partial u^i}(x, u), & i = 1, \ldots, m \\ \\ \frac{dx^i}{dt} = f^i(x, u), & i = 1, \ldots, n. \end{cases} \tag{3.14}$$

We find that the above set of equations (3.14) are exactly the necessary conditions attained by the PMP. Hence we have proved the theorem. □

Relax the assumption that the control space U is open and that the boundary of the set U can be expressed analytically as a function of state, i.e., the set of admissible controls consists of all bounded, piecewise continuous functions $u : [0, 1] \to \mathbb{R}^m$ that satisfy a set of equality and inequality constraints:

$$g_i(t, x, u) = 0, \qquad i = 1, \ldots k, \tag{3.15}$$

$$h_i(t, x, u) \leq 0, \qquad i = 1, \ldots l, \tag{3.16}$$

where $k \leq m$ for the problem to be meaningful.

In the case of inequality constraints we can always turn them into equality constraints by introducing slack variables in the following way:

$$H_i(x,u,w) \equiv h_i(x,u) + w_i^2 = 0, \qquad i = 1,\dots l, \tag{3.17}$$

so that the slack variables $w \in \mathbb{R}^l$ become additional control variables. In this way we have the following set of equality constraints

$$\Omega_i = \begin{cases} F_i(x,u,w) \equiv \dot{x}^i - f^i(x,u) = 0, & i = 1,\dots,n, \\ G_i(x,u,w) \equiv g_i(x,u) = 0, & i = 1,\dots k, \\ H_i(x,u,w) \equiv h_i(t,x,u) + w_i^2 = 0, & i = 1,\dots l, \end{cases} \tag{3.18}$$

The augmented Lagrangian would now be

$$\mathbb{L}(x,\dot{x},u,w,\lambda,\mu,\nu) = L(x,u) + \lambda^i F^i(x,\dot{x},u) + \mu^i G_i(x,u) + \nu^i H_i(x,u,w) \tag{3.19}$$

Now in the case of bounded controls our augmented space x is defined in the following way:

$$\begin{cases} x^i = \mathtt{x}^i, & i = 1,\dots,n, \\ u^i = \dot{\mathtt{x}}^{n+i}, & i = 1,\dots,m, \\ w^i = \dot{\mathtt{x}}^{n+m+i}, & i = 1,\dots,l, \\ \lambda^i = \dot{\mathtt{x}}^{n+m+l+i}, & i = 1,\dots,n, \\ \omega^i = \dot{\mathtt{x}}^{2n+m+l+i}, & i = 1,\dots,k, \\ \mu^i = \dot{\mathtt{x}}^{2n+m+l+k+i}, & i = 1,\dots,l, \end{cases} \tag{3.20}$$

In the new augmented state space, the augmented Lagrangian could be rewritten as

$$\mathbb{L}(\mathtt{x}, \dot{\mathtt{x}}) = L(\mathtt{x}, \mathtt{x}) + \sum_{i=1}^{n+k+l} \dot{\mathtt{x}}^{n+m+l+i}\Omega_i(\mathtt{x}, \dot{\mathtt{x}}) \tag{3.21}$$

Furthermore, at a point t where $\dot{\mathtt{x}}(t)$ has a jump discontinuity, the so called *Weierstrass-Erdmann corner conditions* must hold:

$$\left.\frac{\partial \mathbb{L}}{\partial \dot{\mathtt{x}}}\right|_{t^-}(\mathtt{x}, \dot{\mathtt{x}}) = \left.\frac{\partial \mathbb{L}}{\partial \dot{\mathtt{x}}}\right|_{t^+}(\mathtt{x}, \dot{\mathtt{x}}) \tag{3.22}$$

$$\left.\left(\dot{\mathtt{x}}\frac{\partial \mathbb{L}}{\partial \dot{\mathtt{x}}} - \mathbb{L}\right)\right|_{t^-}(\mathtt{x}, \dot{\mathtt{x}}) = \left.\left(\dot{\mathtt{x}}\frac{\partial \mathbb{L}}{\partial \dot{\mathtt{x}}} - \mathbb{L}\right)\right|_{t^+}(\mathtt{x}, \dot{\mathtt{x}}). \tag{3.23}$$

3.3 Finsler Geometry

Having reduced the optimal control problem into a calculus of variations problem, we shall now introduce Finsler geometry, which was essentially motivated by the Calculus of variations problem. It is characterized by

$$ds = F(x^1, \ldots, x^n\,;\, dx^1, \ldots, dx^n), \tag{3.24}$$

where $F(x\,;\,dx)$, known as the Finsler function, is a smooth, non-negative function in the $2n$ variables, and has the value zero only when $dx = 0$. We now state further conditions on the metric F.

Definition 3.3.1. *Let M be an n-dimensional smooth manifold. A Finsler structure of M is a function*

$$F : TM \to [0, \infty)$$

with the following properties:

1. *Regularity: F is C^∞ on the entire slit tangent bundle $TM \setminus 0$.*

2. *Positive homogeneity: $F(x, \lambda v) = \lambda F(x, v)$ for all $\lambda > 0$ and $v \in T_xM$.*

3. *Strong convexity: the $n \times n$ Hessian matrix*

$$g_{ij} := \left(\frac{1}{2} F^2_{\dot{x}^i \dot{x}^j}\right)$$

   is positive-definite at every point of $TM \setminus 0$.

Given a manifold M and a Finsler structure F on TM, the pair (M, F) is known as the Finsler manifold.

Let $(x^1, \ldots, x^n) = (x^i) : U \to R^n$ be a local coordinate system on an open subset $U \subset M$. As usual, $\{\frac{\partial}{\partial x^i}\}$ and $\{dx^i\}$ are, respectively, the induced coordinate bases for T_xM and T_x^*M.

Definition 3.3.2. *Let M be an n-dimensional smooth manifold. It is said to be a Finsler manifold if the length s of any curve $t \mapsto (x^1(t), \ldots, x^n(t))$, $a \le t \le b$, is given by an integral*

$$s = \int_a^b F\left(x^1, \ldots, x^n; \frac{dx^1}{dt}, \ldots, \frac{dx^n}{dt}\right) dt, \tag{3.25}$$

where the function F has the properties specified in definition (3.3.1).

The positive homogeneity property of the Finsler function F plays a crucial role in the development of the metric. Hence we first state and prove the Euler's lemma for homogeneous functions which we are going to use repeatedly in the sequel.

Theorem 3.3.3. (Euler's Lemma for homogeneous functions): *Suppose a real-valued function H on $\mathbb{R}^n$ is differentiable away from the origin of $\mathbb{R}^n$. Then the following two statements are equivalent:*

- *H is positively homogeneous of degree r. That is,*

$$H(\lambda y) = \lambda^r H(y) \qquad \text{for all } \lambda > 0. \tag{3.26}$$

- *The radial directional derivative of H is r times H. Namely,*

$$y^i H_{y^i}(y) = r\, H(y). \tag{3.27}$$

Proof. Suppose H satisfies $H(\lambda\, y) = \lambda^r H(y)$ for all positive λ. Fix y. Differentiating this equation with respect to the parameter λ gives

$$y^i H_{y^i}(\lambda\, y) = r\lambda^{r-1} H(y). \tag{3.28}$$

Setting λ equal to 1 gives the criterion sought.

Conversely, suppose $y^i H_{y^i}(y) = r\, H(y)$. Fix y and consider the function $H(\lambda\, y)$ with $\lambda > 0$. By the chain rule, we have

$$\frac{d}{d\lambda} H(\lambda\, y) = y^i H_{y^i}(\lambda y) = \frac{1}{\lambda}(\lambda\, y)^i H_{y^i}(\lambda\, y). \tag{3.29}$$

Using our supposition, we see that the last term equals $\frac{1}{\lambda} r H(\lambda y)$. Since we have not assumed that H is nonzero away from the origin, we cannot read the above as $\frac{d}{d\lambda} \log H(\lambda\, y) = \frac{r}{\lambda} = \frac{d}{d\lambda} \log \lambda^r$. Instead, we rewrite it as the ODE

$$\frac{d}{d\lambda} H(\lambda\, y) - \frac{r}{\lambda} H(\lambda\, y) = 0. \tag{3.30}$$

The integrating factor $\frac{1}{\lambda^r}$ then gives $H(\lambda\, y) = C\lambda^r$, where C is some constant that depends on our fixed y. Setting $\lambda = 1$ shows that $C = H(y)$. □

3.3.1 Metric Properties

Let F be a Finsler function on TM. Then, for $(x, \dot{x})$ local coordinates in TM, we have

$$\begin{aligned} F_v(x, \lambda v) &= F_v(x, v), \quad \lambda > 0, \\ F_{vv}(x, \lambda v) &= \frac{1}{\lambda} F_{vv}, \qquad \lambda > 0. \end{aligned}$$

From Euler's theorem on homogeneous functions we then have the following relations of the Finsler function F.

Lemma 3.3.4. *If F is positive homogeneous of degree 1 in $\dot{x}$, then the following* *relations hold*

$$\begin{aligned}\dot{x}^i \cdot F_{\dot{x}^i}(x,\dot{x}) &= F(x,\dot{x}) &(3.31)\\ \dot{x}^j \cdot F_{\dot{x}^i\dot{x}^j}(x,\dot{x}) &= 0 &(3.32)\\ \dot{x}^k \cdot F_{\dot{x}^i\dot{x}^j\dot{x}^k}(x,\dot{x}) &= -F_{\dot{x}^i\dot{x}^j}(x,\dot{x}) &(3.33)\\ \dot{x}^l \cdot F_{\dot{x}^i\dot{x}^j\dot{x}^k\dot{x}^l}(x,\dot{x}) &= -2\,F_{\dot{x}^i\dot{x}^j\dot{x}^k}(x,\dot{x}) &(3.34)\end{aligned}$$

for $(x,\dot{x}) \in TM$.

Proof. Equation (3.31) follows from (3.27) by setting r equal to 1 as the Finsler function F is positive homogeneous of degree 1. Secondly, upon differentiating both sides of (3.31) w.r.t. $\dot{x}^j$, we then have

$$F_{\dot{x}^i\dot{x}^j}(x,\dot{x}) \cdot \dot{x}^i + F_{\dot{x}^j}(x,\dot{x}) = F_{\dot{x}^j}(x,\dot{x})$$

from which equation (3.32) follows

$$F_{\dot{x}^i\dot{x}^j}(x,\dot{x}) \cdot \dot{x}^i = 0, \qquad \forall j = 1,\ldots,n.$$

Similarly, upon differentiating equation (3.32) and (3.33) with respect to $\dot{x}^k$ and $\dot{x}^l$, we end up with equations (3.33) and (3.34), respectively. □

From equation (3.32) we deduce the identity

$$\det|F_{\dot{x}\dot{x}}| = 0. \tag{3.35}$$

Because of the relation (3.35), we are forced to choose our metric as the Hessian of the square of the Finsler function F. We now derive some properties of this metric which will be used later to derive the geodesic equations and other geometric properties.

Lemma 3.3.5. *Given a Finsler function F, the following relations hold true for its square, F^2,*

$$\dot{x}^i \cdot F^2_{\dot{x}^i \dot{x}^j}(x, \dot{x}) = 2\, F(x, \dot{x})\, F_{\dot{x}^j}(x, \dot{x}) \tag{3.36}$$

$$\dot{x}^i \cdot \dot{x}^j \cdot F^2_{\dot{x}^i \dot{x}^j}(x, \dot{x}) = 2\, F^2(x, \dot{x}) \tag{3.37}$$

for $(x, \dot{x}) \in TM$.

Proof. First, we derive the Hessian of the square of the Finsler function F^2, i.e.,

$$\begin{aligned} \frac{1}{2}\, F^2_{\dot{x}^i \dot{x}^j} &= \frac{1}{2}\, (2FF_{\dot{x}^i})_{\dot{x}^j}, \\ &= F_{\dot{x}^i} F_{\dot{x}^j} + F F_{\dot{x}^i \dot{x}^j}. \end{aligned} \tag{3.38}$$

Multiplying equation (3.38) with $\dot{x}^i$ on both sides and using equations (3.31) and (3.32) we prove relation (3.36)

$$\begin{aligned} \frac{1}{2}\, F^2_{\dot{x}^i \dot{x}^j} \cdot \dot{x}^i &= (F_{\dot{x}^i} F_{\dot{x}^j} + F F_{\dot{x}^i \dot{x}^j}) \cdot \dot{x}^i \\ &= F_{\dot{x}^j}(F_{\dot{x}^i} \cdot \dot{x}^i) + F(F_{\dot{x}^i \dot{x}^j} \cdot \dot{x}^i) \\ &= F F_{\dot{x}^j}. \end{aligned} \tag{3.39}$$

Upon multiplying equation (3.36) with $\dot{x}^j$ on both sides and using equations (3.39) and (3.31) we have the following

$$\begin{aligned}\frac{1}{2}F^2_{\dot{x}^i\dot{x}^j}\dot{x}^i\dot{x}^j &= \dot{x}^i\left(\frac{1}{2}F^2_{\dot{x}^i\dot{x}^j}\dot{x}^j\right)\\ &= \dot{x}^i F F_{\dot{x}^i}\\ &= F(F_{\dot{x}^i}\dot{x}^i)\\ &= F^2 \end{aligned} \tag{3.40}$$

□

Equation (3.37) suggests us that the metric g_{ij} should indeed be of the form

$$g_{ij} = \frac{1}{2}\frac{\partial F^2}{\partial \dot{x}^i \partial \dot{x}^j}(x, \dot{x}) \tag{3.41}$$

which is a consequence of the Finsler function F being positive homogeneous of degree 1. Furthermore, from equation (3.37), it follows that

$$g_{ij}(x, \dot{x})\dot{x}^i\dot{x}^j = F^2. \tag{3.42}$$

Since the metric $g_{ij}(x, \dot{x})$ is positive homogeneous of degree zero in $\dot{x}$, we have

$$g_{ij}(x, v)v^i v^j = 1, \tag{3.43}$$

where $v^i = \frac{\dot{x}^i}{F}$ are unit vectors.

Theorem 3.3.6. *Let F be a non-negative real-valued function on $\mathbb{R}^n$, then we have the following conclusions:*

- (Positivity)

$$F(x, \dot{x}) > 0, \qquad \textit{whenever } \dot{x} \neq 0. \tag{3.44}$$

- (Fundamental inequality)

$$w^i \cdot F_{v^i}(x, v) \leq F(x, w) \qquad \textit{at all } v \neq 0, \tag{3.45}$$

and equality holds if and only if $w = \alpha\, v$ *for some* $\alpha \geq 0$.

Proof. Positivity: Consider equation (3.41), the hypothesized strong convexity of F says that the left-hand side is positive whenever $\dot{x} \neq 0$, thus F is strictly positive on $\mathbb{R}^n \setminus 0$.

Fundamental inequality: The consequences of Euler's theorem are used repeatedly here. When $w = \alpha\, \dot{v}$ for some $\alpha \geq 0$, both sides equal $\alpha F(x, v)$. When w is a negative multiple of v, the inequality is strict because its left-hand side becomes a negative multiple of $F(x, v)$. The case $w \neq 0$ but $v = 0$ is disallowed. Lastly, suppose v, w are linearly independent, then so are v and $\xi := v - w$. Inequality (3.45) now reads

$$\begin{aligned} F(x, w) &> F(x, v) - F_{v^i}(x, v)(v^i - w^i), \\ &> w^i \cdot F_{v^i}(x, v) \end{aligned}$$

which readily reduces to the strict part of (3.45). We thus have completed the proof. □

Theorem 3.3.7. (Positive-Definiteness) *Suppose $F(x,v)$ is a Finsler function. Then $F_{v^iv^j}$ is positive semi-definite, i.e.,*

$$\begin{aligned} F_{v^iv^j}\cdot w^i\cdot w^j &\geq 0, && \forall\, w\in\mathbb{R}^n, \\ F_{v^iv^j}\cdot w^i\cdot w^j &= 0, && \textit{if}\ \ w=\lambda v,\ \ \lambda>0. \end{aligned}$$

Proof. At each point $v\in\mathbb{R}^n\setminus 0$, the matrix g_{ij} defines an inner product. So we have the Cauchy-Schwarz type inequality

$$\big[\,g_{ij}(x,v)w^iu^j\,\big]^2 \;\leq\; \big[\,g_{ij}(x,v)w^iw^j\,\big]\cdot\big[\,g_{ij}(x,v)u^iu^j\,\big] \qquad \forall\, w,u\in\mathbb{R}^n, \quad (3.46)$$

where equality holds if and only if w and u are collinear. Setting $u^i=v^i$ and using (3.42), we obtain

$$\big[\,g_{ij}(x,v)w^iv^j\,\big]^2 \;\leq\; F^2(x,v)\cdot\big[\,g_{ij}(x,v)w^iw^j\,\big] \qquad \forall\, w\in\mathbb{R}^n, \quad (3.47)$$

where equality holds if and only if w and v are collinear. On the other hand, the formula (3.41) for g_{ij} leads us to

$$F_{v^iv^j}(x,v)w^iw^j \;=\; \frac{1}{F^3(x,v)}\,\Big\{F^2(x,v)\,\big[\,g_{ij}(x,v)w^iw^j\,\big]-\big[\,g_{ij}(x,v)v^iw^j\,\big]^2\Big\} \quad (3.48)$$

which, in conjunction with (3.47), gives

$$F_{v^iv^j}(x,v)w^iw^j \;\geq\; 0 \qquad \forall\, w\in\mathbb{R}^n. \quad (3.49)$$

In the above equation, equality holds if and only if v and w are collinear. ∎

Theorem 3.3.8. (Triangle inequality) *If* $v, w \in T_xM$, *then*

$$F(x, v + w) \leq F(x, v) + F(x, w). \tag{3.50}$$

Proof. Suppose $u, v \in T_xM$ are two non-collinear vectors, then by the second mean-value theorem, we have

$$F(x, u) = F(x, v) + F_{v^i}(x, v) \cdot (u^i - v^i) + \frac{1}{2} F_{v^i v^j}(x, \theta v + (1 - \theta)u) \cdot (u^i - v^i) \cdot (u^j - v^j). \tag{3.51}$$

for some $0 < \theta < 1$.By applying equation (3.31) to the second term of RHS we get

$$F(x, u) = F_{v^i}(x, v) \cdot u^i + \frac{1}{2} F_{vv}(x, \theta v + (1 - \theta)u) \cdot (u^i - v^i) \cdot (u^j - v^j). \tag{3.52}$$

Since u, v are linearly independent, by the fundamental inequality (3.43) we deduce that

$$F(x, u) \geq F_{v^i}(x, v) \cdot u^i. \tag{3.53}$$

Let $v = u + w$, where $u, w \in T_xM$. Then

$$F_{v^i}(x, v + w) \cdot v^i \quad \leq \quad F(x, v) \tag{3.54}$$

$$F_{v^i}(x, v + w) \cdot w^i \quad \leq \quad F(x, w). \tag{3.55}$$

Summing these two inequalities we get

$$F_{v^i}(x, v + w) \cdot (v^i + w^i) \leq F(x, v) + F(x, w). \tag{3.56}$$

Substituting equations (3.54) and (3.55) in (3.56), we have

$$F(x, v + w) \leq F(x, v) + F(x, w).$$

□

3.3.2 Geometry of Finsler Manifolds

In this section we shall outline the geometry of Finsler manifolds and introduce metrics for the tangent and cotangent spaces. At each point x of the manifold M we consider the tangent spaces T_xM and T^*_xM, of the contravariant and covariant vectors at x respectively. Corresponding to each contravariant vector $v^i \in T_xM$ there is defined a covariant vector $y_i \in T^*_xM$ and vice versa. A concrete picture of this mapping can be given in terms of certain hypersurfaces in the two tangent spaces.

Before we introduce these hypersurfaces it is convenient to assume the following condition on the Finsler function F:

$$F(x^i, v) > 0, \qquad \text{for all } (x, v) \in TM \setminus 0.$$

This condition holds true only for this section in order to present a clear geometric picture.

Indicatrix

We shall now derive an appropriate metric on the tangent space TM of the Finsler manifold M with a given Finsler function $F(x, v)$.

Definition 3.3.9. *The indicatrix is an $(n-1)$-dimensional hypersurface in T_xM defined by*

$$F(x, v) = 1, \qquad \textit{for all} \quad v \in T_xM. \tag{3.57}$$

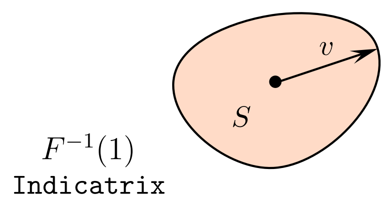

Figure 3.2: Indicatrix

Theorem 3.3.10. *The set $S = \{v \in T_xM \mid F(x, v) \leq 1\}$ is convex.*

Proof. Let $u \in T_xM$. Then λu intersects the hyper-surface S in one and only one point. By positive definiteness property $F(x, v) \geq 0$ and by homogeneity every vector λu for an appropriate λ will intersect the hypersurface at one point.

To prove the above assertion we assume that vectors λv and μv intersect the surface S, then

$$\begin{aligned} F(x, \lambda v) = \lambda F(x, v) &= 1, \\ F(x, \mu v) = \mu F(x, v) &= 1. \end{aligned}$$

From which it follows that, $0 = (\lambda - \mu)F(x, v) = F(x, (\lambda - \mu)v)$ which contradicts (3.57).

Now let u, w be such that $F(x, u) \leq 1$ and $F(x, w) \leq 1$. Let $v = \theta u + (1 - \theta)w$ with $\theta \in [0, 1]$. Then by applying Theorem (3.3.8),

$$F(x, v) \leq F(x, \theta u) + F(x, (1 - \theta)w), \tag{3.58}$$

and by homogeneity

$$F(x, v) \leq \theta F(x, u) + (1 - \theta)F(x, w). \tag{3.59}$$

It follows that

$$F(x, v) \leq 1.$$

Therefore, $v \in S$. Since u and w are two arbitrary points of S, the result follows.

□

Metric Tensor

It is to be noted that, the metric

$$ds = F(x, v), \qquad v \in T_x M \tag{3.60}$$

on the Finsler manifold M gives rise to a metric or measurement of length in each tangent space in a unique manner. The length or norm of a tangent vector $v \in T_x M$ is given by

$$|v| = F(x, v). \tag{3.61}$$

Geometrically, this is simply the ratio OQ/OR where R is the point at which OQ, elongated if necessary, intersects the indicatrix and O is the origin of the tangent space T_xM. The length of a vector joining two points $v,\ w \in T_xM$, is given by $F(x, v - w)$. It is clear therefore, that the indicatrix plays the role of the unit sphere in the geometry of the tangent vector space T_xM. As defined earlier we

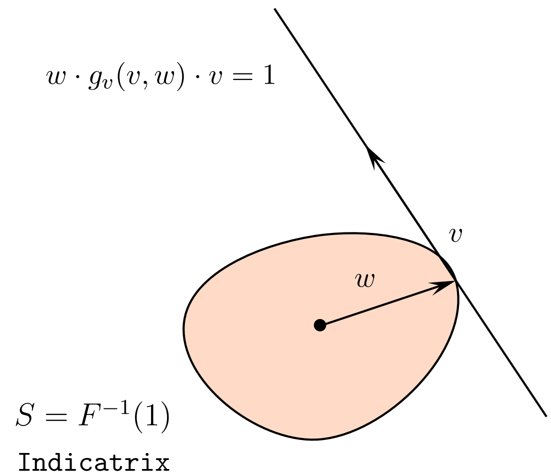

Figure 3.3: Hyperplane to Indicatrix

have the quantities g_{ij} as

$$g_{ij}(x, v) = \frac{1}{2}\frac{\partial^2 F^2}{\partial v^i \partial v^j}(x, v). \tag{3.62}$$

In view of the homogeneity condition, the function F^2 is positively homogeneous of second degree in tangent vector v, and hence we have

$$F^2(x, v) = g_{ij}(x, v)v^i v^j. \tag{3.63}$$

So all lengths in the tangent space T_xM may now be expressed in terms of the g_{ij}. The equation of the indicatrix of T_xM may now be written in the form:

$$F^2(x, v) = g_{ij}(x, v)v^i v^j = 1. \tag{3.64}$$

A tangent hyperplane to the indicatrix at a point (x, w) for $w \in T_xM$ satisfies the following equation

$$F_v(x, w)(v - w) = 0, \qquad \text{for all} \quad v \in T_xM \tag{3.65}$$

with $F(x, w) = 1$.

Proposition 3.3.11. *The hyperplane equation (3.65) is equivalent to*

$$w^T \cdot g(x, w) \cdot v = 1. \tag{3.66}$$

Proof. Expanding equation (3.65), we get

$$F_v(x, w)v = F_v(x, w)w. \tag{3.67}$$

But from equations (3.36) and (3.57), we conclude that

$$F_v(x, w)v = 1. \tag{3.68}$$

Using (3.36) we have

$$w \cdot \frac{1}{2} F^2_{vv}(x, w) = F_v(x, w) \tag{3.69}$$

Post-multiplying equation (3.69) by v and using (3.31) we have

$$w \cdot F^2_{vv}(x, w) \cdot v = F_v(x, w) \cdot v = 1 \tag{3.70}$$

□

Having thus defined a metric in each tangent space T_xM of the Finsler manifold M, we may conversely assert that the metric in a tangent space T_xM determines the local metric of M in the immediate vicinity of the point $x \in M$. For strictly speaking, the expression $F(x, v)$ represents the length of a vector v of the tangent space T_xM, which we can interpret as an element ds of length on M, namely as a first-order approximation.

Fundamental Inequality

The fundamental inequality could be viewed as an extension of the Euler's theorem, from an equation to an inequality, i.e., the fundamental inequality

$$w^i F_{v^i}(x, v) \leq w, \tag{3.71}$$

becomes the Euler's theorem (3.31) for the metric F when $w = \alpha\, v$ for some $\alpha \geq 0$. Furthermore, by adding equation (3.31) to the fundamental inequality (3.43), we have

$$F(x, v) + F_{v^i}(x, v) \cdot (w^i - v^i) \;\leq\; F(x, w), \tag{3.72}$$

where equality holds only when $w = \alpha\, v$ with $\alpha \geq 0$. If one thinks of v as being fixed and w as the independent variable, then the left-hand side is the linear approximation of the value of $F(w)$. So, at any fixed $(v, F(x, v))$ on the graph of F, the tangent hyperplane touches the graph only along the ray $(\alpha\, v, \alpha\, F(x, v))$, $\alpha \geq 0$. Everywhere else, the tangent hyperplane lies below the graph of F. This is depicted in Figure 3.4. Thus the graph of F is a convex cone at the origin of our

Minkowski space. Since $F(x,v) > 0$ for $v \neq 0$, we can multiply the fundamental

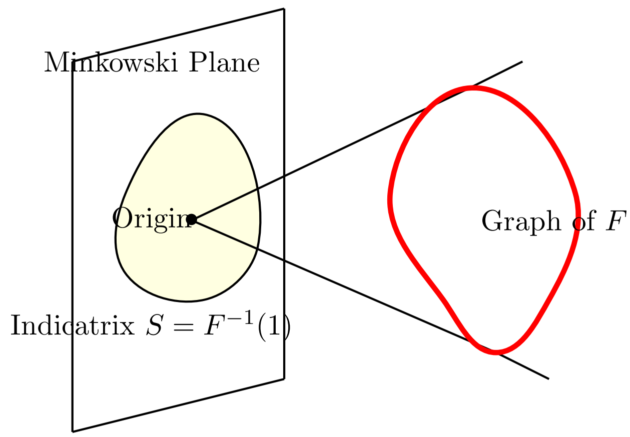

Figure 3.4: Indicatrix

inequality by $F(x,v)$ to get

$$w^i\, F(x,v)\, F_{v^i}(x,v) \;\leq\; F(x,v)\, F(x,w). \tag{3.73}$$

Now substituting the relation $v^i g_{ij}(x,v) \;=\; F\, F_{v^i}$ in equation (3.73), we have

$$g_{ij}(x,v)\; w^i\, v^j \;\leq F(x,w)\, F(x,v). \tag{3.74}$$

One can view equation (3.74) as a generalization of the Cauchy-Schwarz inequality, from inner products to Minkowski norms. When written in the metric form, we have

$$\left[\, g_{ij}(x,v)\, w^i\, v^j \,\right]^2 \;\leq\; \left[\, g_{pq}(x,w)\, w^p\, w^q \,\right] \cdot \left[\, g_{rs}(x,v)\, v^r\, v^s \,\right]. \tag{3.75}$$

Figuratrix

In the Finsler geometry the momenta variables p_i cannot be defined in the usual way in classical mechanics, i.e.,

$$p_i = \frac{\partial F(x, \dot{x})}{\partial \dot{x}}, \tag{3.76}$$

because of the homogeneity property we cannot express the velocity vectors $\dot{x}$ as functions of (x^i, p_i). We now show the corresponding canonical variables y_i and the Hamiltonian function $H(x^i, y_i)$ for the Finsler function $F(x^i, \dot{x}^i)$.

We now define alternative quantities y_i by the relation

$$y_i = g_{ij}(x^k, \dot{x}^k)\dot{x}^j, \tag{3.77}$$

which play the role of the canonical momenta in the Finslerian case. Using the definition of the metric tensor g_{ij}, we can write the momenta variables as

$$y_i = F\,\frac{\partial F}{\partial \dot{x}^i}. \tag{3.78}$$

Thus one can see that the momenta variables p_i in the nonhomogeneous case and the momenta variables y_i in the Finsler case differ by a multiplicative factor F. The relations (3.77) can be solved uniquely for the velocity variables $\dot{x}^i$ as functions of (x^i, y_i) as a result of the condition

$$\det g_{ij}(x^k, \dot{x}^k) \neq 0. \tag{3.79}$$

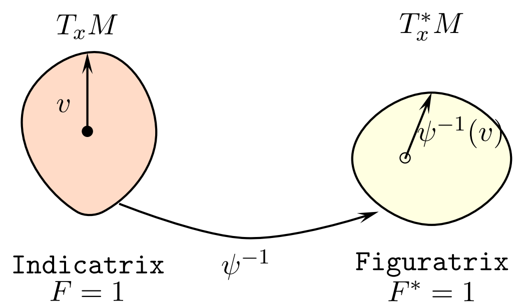

Figure 3.5: Legendre Transform

By differentiating the momenta y_i with respect to the velocity $\dot{x}^k$, we have

$$\frac{\partial y_i}{\partial \dot{x}^k} = \frac{\partial g_{ij}}{\partial \dot{x}^k}\dot{x}^j + g_{ij}\delta^j_k$$

$$= g_{ik}. \tag{3.80}$$

Thus the Jacobian of the momenta is simply the matrix g_{ij}, and hence we may write

$$\dot{x}^i = \psi^i(x^j, y_j). \tag{3.81}$$

Furthermore, the metric g_{ij} possesses an inverse g^{ij}. In the later we substitute from (3.81), thus obtaining a matrix $g^{ij}(x, y)$, such that

$$g^{ij}(x^k, y_k)\; g_{ih}(x^k, \dot{x}^k) = \delta^j_h, \tag{3.82}$$

whenever $\dot{x}^k$ and y_k are related by equation (3.77).

It follows from (3.78) that the momenta y_i are positively homogeneous of the first degree in $\dot{x}^j$. The converse must necessarily hold for the inverse equation (3.81). Thus, since (3.80) may now be used to solve the equations (3.77) explicitly for $\dot{x}^i$ in the form

$$\dot{x}^i = g^{ij}(x, y)\, y_j, \tag{3.83}$$

where the g^{ij} are positively homogeneous of degree zero in the momenta y.

The identity (3.82) suggests that we should define the Hamiltonian function $H(x, y)$ by the relation

$$H^2(x, y) = g^{ij}(x, y) y_i y_j. \tag{3.84}$$

We shall now give justification for this form of the Hamiltonian but later we shall use the Legendre transformation to actually derive this form.

Firstly, on differentiating (3.82) with respect to $\dot{x}^l$ and applying (3.80), we obtain

$$\frac{\partial g^{ij}}{\partial y_k} y_i = 0. \tag{3.85}$$

Secondly, if we substitute from (3.77) for y_i and y_j in (3.84) we deduce from (3.82) that

$$H^2(x, y) = g_{ij}(x, \dot{x}) \dot{x}^i \dot{x}^j = F^2(x, \dot{x}). \tag{3.86}$$

Strictly speaking, equation (3.86) should be expressed as

$$H(x, y) = F(x, \psi(x, y)), \tag{3.87}$$

where we have replaced $\dot{x}$ by (3.81). Equation (3.77) together with its inverse (3.83), represents a continuous 1-1 mapping of the two tangent spaces T_xM and T_x^*M.

3.4 Optimal Control Geodesics

3.4.1 Weierstrass's side condition

Weierstrass considered the problems of Calculus of Variations with Lagrangians L such that $L(\mathtt{x}, \dot{\mathtt{x}})$ is positively homogeneous with respect to the velocity $\dot{\mathtt{x}}$ and does not depend on time, i.e.,

$$L(\mathtt{x}, \alpha\dot{\mathtt{x}}) = \alpha L(\mathtt{x}, \dot{\mathtt{x}}), \qquad \text{for all } (\mathtt{x}, \dot{\mathtt{x}}) \in TM, \text{ and } \alpha > 0. \tag{3.88}$$

In order to use the Finsler approach we need our Lagrangians L to be positive homogeneous of degree one. However the Lagrangians for the calculus of variations problems are not in general homogeneous in $\dot{\mathtt{x}}$, the following theorem gives us a method of forming Lagrangians L which are positive homogeneous in $\dot{\mathtt{x}}$ of degree 1.

Let γ be a parameterized curve in the manifold M:

$$\gamma : t \in [0, 1] \to \gamma(t) \in M.$$

The curve γ has an analytical expression of the form:

$$\gamma \equiv \mathtt{x}(t) = (\mathtt{x}^1(t), \ldots, \mathtt{x}^n(t)), \qquad t \in [0, 1].$$

The Tangent lift of the curve γ to the tangent space TM of the Finsler manifold M is given as

$$(\gamma(t), \dot{\gamma}(t)) = (\mathtt{x}^i(t), \dot{\mathtt{x}}^i(t)), \qquad t \in [0,1], \;\; i = 1, \dots n.$$

In homogenizing the Lagrangian $L(\mathtt{x}, \dot{\mathtt{x}})$, we move from our configuration space $\mathtt{x} = (\mathtt{x}^1, \dots, \mathtt{x}^n)$ to the space of events $\mathrm{x} = (t, \mathtt{x}) = (\mathrm{x}^0, \mathrm{x}^1, \dots, \mathrm{x}^n)$, where $\mathrm{x}^0 = t$ and $\mathrm{x}^i = \mathtt{x}^i$ for $i = 1, \dots n$. The dimension of our new event space would then be $n+1$. We now parameterize the curve $\gamma(t)$ with a new parameter τ by the following transformation function σ: $\tau = \sigma(t)$, and its inverse $\rho \equiv \sigma^{-1}$ exists such that σ and ρ are 1-1 differentiable monotone maps in the domain $[0,1]$.

The tangent lift of a curve in the space of events $(\mathrm{x}^i, \mathrm{x}'^i) := (t, \mathtt{x}^\alpha, t', \mathtt{x}'^\alpha)$, where $t' := \frac{d\rho}{d\tau}(\sigma(t))$, $x'^\alpha := \frac{dx^\alpha}{d\tau}(\sigma(t))$, $i = 0, 1 \dots, n$ and $\alpha = 1, \dots, n$. Here τ can be considered as "pseudo-time".

Let us now introduce a new Lagrangian function Λ such that

$$\begin{aligned} \Lambda &\equiv \Lambda\left(\rho(\sigma(t)), x(\sigma(t)), \frac{d\rho}{d\tau}(\sigma(t)), \frac{dx}{d\tau}(\sigma(t))\right) \\ &= L\left(\rho(\sigma(t)), x(\sigma(t)), \frac{dx}{d\tau}(\sigma(t)) \cdot \left(\frac{d\rho}{d\tau}(\sigma(t))\right)^{-1}\right) \frac{d\rho}{d\tau}(\sigma(t)) \end{aligned} \tag{3.89}$$

The following theorem shows that the Lagrangian Λ is positive homogeneous of degree 1 in x'.

Theorem 3.4.1. *Let τ be an arbitrary parameter and the Lagrangian, $L(t, \mathtt{x}(t), \dot{\mathtt{x}}(t))$, not necessarily positive homogeneous of degree 1 in $\dot{\mathtt{x}}$, then the Lagrangian Λ defined by $\Lambda(\mathrm{x}, \mathrm{x}') \equiv L(t, \mathtt{x}, \mathtt{x}'/t')\, t'$ is positive homogeneous of degree 1 in x'.*

Proof. For $\alpha > 0$ we have,

$$\begin{aligned}
\Lambda(\mathrm{x}, \alpha\mathrm{x}') &= \Lambda(t, \mathtt{x}, t', \mathtt{x}') \\
&= L(\mathtt{x}, \frac{\alpha\mathtt{x}'}{\alpha t'}) \cdot \alpha t' \\
&= \alpha\, L(\mathtt{x}, \frac{\mathtt{x}'}{t'}) t' \\
&= \alpha\, \Lambda(\tau, \mathtt{x}, t', \mathtt{x}') \equiv \alpha\, \Lambda(\mathrm{x}, \mathrm{x}')
\end{aligned}$$

Hence the new Lagrangian, Λ, is positive homogeneous of degree 1 in x'. □

Theorem 3.4.2. *The following minimization problems are equivalent:*

$$\arg \min_{\mathtt{x}} \int_0^1 L(\mathtt{x}, \dot{\mathtt{x}}) dt = \arg \min_{\mathtt{x}} \int_{\sigma(0)}^{\sigma(1)} \Lambda(\mathrm{x}, \mathrm{x}') d\tau,$$

for $x(0) = x_0,\ x(1) = x_1$.

Proof.

$$\begin{aligned}
\int_0^1 L(\mathtt{x}, \dot{\mathtt{x}}) dt &= \int_0^1 L(t, \mathtt{x}, \frac{d\mathtt{x}}{dt}) dt \\
&= \int_0^1 L\left(t, \mathtt{x}, \left(\frac{dt}{d\tau}\right)^{-1} \frac{d\mathtt{x}}{d\tau}\right) \frac{dt}{d\tau} dt \\
&= \int_{\sigma(0)}^{\sigma(1)} L\left(t, \mathtt{x}, \left(\frac{dt}{d\tau}\right)^{-1} \frac{d\mathtt{x}}{d\tau}\right) d\tau \\
&= \int_{\sigma(0)}^{\sigma(1)} \Lambda(t, \mathtt{x}, t', \mathtt{x}') d\tau
\end{aligned}$$

□

So far we have left the parameter τ arbitrary as long as there exists a diffeomorphism between the parameter τ and time t. From now on we shall interpret our homogenizing parameter τ to be arc-length, i.e., $\tau = s$, as we show below that there exists a diffeomorphism between arc-length s and time t. Indeed, the arc-length function $s = \sigma(t)$, $t \in [0,1]$, given by

$$s(t) = \int_0^t L\left(x(\tau), \frac{dx}{dt}(\tau)\right) d\tau, \qquad t \in [0,1], \tag{3.90}$$

is derivable, having the derivative:

$$\frac{ds}{dt} = L\left(x(\tau), \frac{dx}{dt}(\tau)\right) > 0, \qquad t \in (0,1), \tag{3.91}$$

since we consider only positive Lagrangians in its domain. So the function $s = \sigma(t)$, $t \in [0,1]$, is invertible and also differentiable by the fundamental theorem of calculus. Let $t = \rho(s)$ be its inverse. Then we have $\rho(\sigma(t)) = t$, which upon differentiating with respect to t, we have

$$\frac{d\rho}{ds}(\sigma(t)) \cdot \frac{d\sigma}{dt}(t) = 1. \tag{3.92}$$

Therefore, it follows that

$$\frac{d\rho}{ds}(\sigma(t)) = \frac{1}{\frac{d\sigma}{dt}(t)} > 0, \tag{3.93}$$

The change of parameter $t \to s$, given by $s = s(t)$, has the property

$$F\left(x(s), \frac{dx}{ds}(s)\right) = 1. \tag{3.94}$$

Our new arc-length element ds is therefore given as

$$ds = \Lambda\left(\mathrm{x}^i, \frac{d\mathrm{x}^i}{d\tau}\right) d\tau, \qquad i = 0, 1, \ldots, n, \tag{3.95}$$

where

$$\Lambda\left(\mathrm{x}^i, \frac{d\mathrm{x}^i}{d\tau}\right) = L\left(\mathrm{x}^i, \frac{d\mathrm{x}^\alpha}{d\mathrm{x}^0}\right) \frac{d\mathrm{x}^0}{d\tau} \qquad \alpha = 1, \ldots, n. \tag{3.96}$$

and by interpreting the parameter τ to be arc-length s, we have

$$\frac{ds}{d\tau} = \Lambda\left(\mathrm{x}^i, \frac{d\mathrm{x}^i}{d\tau}\right) \equiv 1, \tag{3.97}$$

and from equation (3.96) it follows that

$$\left(\frac{dx^0}{ds}\right)^{-1} = \left(\frac{dt}{ds}\right)^{-1} = L\left(\mathrm{x}^i, \frac{d\mathrm{x}^\alpha}{dt}\right). \tag{3.98}$$

Upon differentiating equation (3.97) with respect to arc-length s, we have

$$\frac{d\Lambda}{ds}\left(\mathrm{x}^i, \frac{d\mathrm{x}^i}{ds}\right) = 0. \tag{3.99}$$

Theorem 3.4.3. (Parameter Independence of Length) *Let $\gamma : [a, b] \to M$ be a piecewise smooth curve segment defined in the closed interval $[a, b]$, where M is a Finsler manifold. Let $\sigma : [\alpha, \beta] \to [a, b]$ be a diffeomorphism. Then the curves γ and $\gamma \circ \sigma : [\alpha, \beta] \to M$ have the same arc-length; namely, the following holds:*

$$I(\gamma) = I(\gamma \circ \sigma),$$

where

$$I(\gamma) = \int_a^b \Lambda(\gamma(\tau), \dot{\gamma}(\tau))d\tau.$$

Proof. We use the fact that σ is a differentiable monotone map, i.e.,

$$\frac{d\sigma}{ds} > 0.$$

We then have

$$\begin{aligned}
I(\tilde{\gamma}) &= \int_\alpha^\beta \Lambda(\tilde{\gamma}(\tau), \dot{\tilde{\gamma}}(\tau))d\tau \\
&= \int_\alpha^\beta \Lambda\left((\gamma \circ \sigma)(\tau), \frac{d(\gamma \circ \sigma)}{d\tau}(\tau)\right) d\tau \\
&= \int_\alpha^\beta \Lambda\left(\gamma(\sigma(\tau)), \frac{d\gamma}{d\sigma}(\sigma(\tau)) \cdot \frac{d\sigma}{d\tau}(\tau)\right) d\tau \\
&= \int_\alpha^\beta \Lambda\left(\gamma(\sigma(\tau), \frac{d\gamma}{d\sigma}(\sigma(\tau))\right) d\tau \cdot \frac{d\sigma}{d\tau} \\
&= \int_a^b \Lambda\left(\gamma(\sigma), \frac{d\gamma}{d\sigma}(\sigma)\right) d\sigma \\
&= I(\gamma).
\end{aligned}$$

□

As we have seen in the previous section that one defines a Finsler metric for the optimal control problem by first defining a new coordinate system $\mathrm{x} = \mathrm{x}(s) \in$

$M \subseteq \mathbb{R}^{2n+m+1}$ parameterized by arc-length and an element of the tangent bundle of M is given by $(\mathrm{x}, \mathrm{x}') = (\mathrm{x}, (\mathrm{t}'(s), \mathrm{y}'(s), \mathrm{c}(s), \mathrm{p}(s))) \in TM \subseteq \mathbb{R}^{4n+2m+2}$. Given a Lagrangian and the control system for the optimal control problem, the metric is defined as

$$g_{ij}(\mathrm{x}, \mathrm{x}') = \frac{1}{2}\frac{\partial^2 F}{\partial \mathrm{x}'_i \partial \mathrm{x}'_j}(\mathrm{x}, \mathrm{x}'), \qquad i, j = 1, \ldots, 2n+m+1 \quad (3.100)$$

where

$$F(\mathrm{x}, \mathrm{x}') = \tilde{L}(\mathrm{x}, \mathrm{x}')^2 \quad (3.101)$$

$$\tilde{L}(\mathrm{x}, \mathrm{x}') = L\left(\mathrm{y}, \frac{\mathrm{c}}{\mathrm{t}'}\right) + \frac{\mathrm{p}}{\mathrm{t}'} \cdot \left(\frac{\mathrm{y}'}{\mathrm{t}'} - f\left(\mathrm{y}, \frac{\mathrm{c}}{\mathrm{t}'}\right)\right). \quad (3.102)$$

The metric g_{ij} which is a symmetric matrix of dimension $2n+m+1 \times 2n+m+1$ is given as follows

$$G(\mathrm{x}, \dot{\mathrm{x}}) = \begin{bmatrix} [F_{\mathrm{t}'\mathrm{t}'}]_{1\times 1} & [F_{\mathrm{t}'\mathrm{y}'^i}]_{1\times n} & [F_{\mathrm{t}'\mathrm{c}^i}]_{1\times m} & [F_{\mathrm{t}'\mathrm{p}^i}]_{1\times n} \\ [F_{\mathrm{y}'^i\mathrm{t}'}]_{n\times 1} & [F_{\mathrm{y}'^i\mathrm{y}'^j}]_{n\times n} & [F_{\mathrm{y}'^i\mathrm{u}^j}]_{n\times m} & [F_{\mathrm{y}'^i\mathrm{p}^j}]_{n\times n} \\ [F_{\mathrm{c}^i\mathrm{t}'}]_{m\times 1} & [F_{\mathrm{c}^i\mathrm{y}'^i}]_{m\times n} & [F_{\mathrm{c}^i\mathrm{c}^j}]_{m\times m} & [F_{\mathrm{c}^i\mathrm{p}^j}]_{m\times n} \\ [F_{\mathrm{p}^i\mathrm{t}'}]_{n\times 1} & [F_{\mathrm{p}^i\mathrm{y}'^j}]_{n\times n} & [F_{\mathrm{p}^i\mathrm{c}^j}]_{n\times m} & [F_{\mathrm{p}^i\mathrm{p}^j}]_{n\times n} \end{bmatrix} \quad (3.103)$$

3.4.2 Finsler Geodesics

We shall now derive the Euler-Lagrange equations for locally minimizing curves in a Finsler space (M, F). Let $c : [a, b] \to M$ be a constant speed piecewise C^∞ curve $F(c, \dot{c}) = \lambda = constant$. By definition, there is a partition of $[a, b]$,

$$a = t_0 < \cdots < t_k = b,$$

such that c on each $[t_{i-1}, t_i]$ is C^∞. Fix the above partition and consider a piecewise C^∞ map $H : (-\epsilon, \epsilon) \times [a, b] \to M$ such that

1. H is C^0 on $(-\epsilon, \epsilon) \times [a, b]$;
2. H is C^∞ on each $(-\epsilon, \epsilon) \times [t_{i-1}, t_i]$, $i = 1, \ldots, k$;
3. $c(t) = H(0, t)$, $a \le t \le b$.

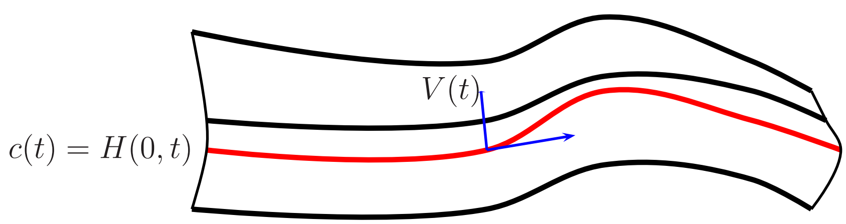

Figure 3.6: Variation

Then the vector field

$$V(t) = V^i(t) \frac{\partial}{\partial x^i}\bigg|_{c(t)} := \frac{\partial H}{\partial u}(0, t)$$

is called the variation field of H. The length of $c_u(t) := H(u,t)$ is given by

$$L(u) \ := \ \int_a^b F(c_u(t), \dot{c}_u(t))dt \ = \ \sum_{i=1}^{k} \int_{t_{i-1}}^{t_i} F\left(H(u,t), \frac{\partial H}{\partial t}(u,t) \right) dt$$

Since the first variation should vanish for the curve $c(t)$ to be an extremal we now compute

$$\begin{aligned}
L'(0) \ &= \ \int_a^b \frac{1}{2F} \left\{ F^2_{x^k} V^k + F^2_{\dot{x}^k} \frac{dV^k}{dt} \right\} dt \\
&= \ \int_a^b \left\{ \frac{1}{2F} F^2_{x^k} V^k - \frac{d}{dt} (\frac{1}{2F} F^2_{\dot{x}^k}) \right\} V^k dt + \sum_{i=1}^{k} \frac{1}{2F} F^2_{\dot{x}^k} V^k \Big|_{t_{i-1}}^{t_i} \\
&= \ \int_a^b \frac{1}{2F} \left\{ F^2_{x^k} - F^2_{x^l \dot{x}^k} \dot{c}^l - F^2_{\dot{x}^k \dot{x}^l} \ddot{c}^l \right\} V^k dt + \sum_{i=1}^{k} \frac{1}{F} g_{jk} \dot{c}^j V^k \Big|_{t_{i-1}}^{t_i} \\
&= \ - \int_a^b \frac{1}{F} g_{jk} \left\{ \ddot{c}^l + 2G^j(c, \dot{c}) \right\} V^k dt + \sum_{i=1}^{k} \frac{1}{F} g_{jk} \dot{c}^j V^k \Big|_{t_{i-1}}^{t_i} \qquad (3.104)
\end{aligned}$$

where

$$G^i(x,\dot{x}) \ = \ \frac{1}{4} g^{il} \ \left[F^2_{x^k \dot{x}^l} \, \dot{x}^k - F^2_{x^l} \right]. \qquad (3.105)$$

The geodesic curvature of the geodesic c at $c(t)$, $\kappa(t)$, is given by

$$\kappa(t) \ := \ \frac{1}{F^2} \left\{ \ddot{c}^i + 2G^i(c,\dot{c}) \right\} \ \frac{\partial}{\partial x^i} \Big|_{c(t)}. \qquad (3.106)$$

We can now express the first variation in index-free form as follows:

$$
\begin{aligned}
L'(0) \;=\; & -\lambda \int_a^b g_{\dot c}(\kappa, V)dt + \frac{1}{\lambda} g_{\dot c(b)}(\dot c(b), V(b)) - g_{\dot c(a)}(\dot c(a), V(a)) \\
& + \frac{1}{\lambda} \sum_{i=1}^{k} \Big\{ g_{\dot c(t_i^-)}(\dot c(t_i^-), V(t_i) - g_{\dot c(t_i^+)}(\dot c(t_i^+), V(t_i)) \Big\}
\end{aligned}
\tag{3.107}
$$

where $\lambda = F(c, \dot c)$ is a constant by assumption.

Assuming that c has minimum length, it then follows that

$$L'(0) = 0$$

for any piecewise cure C^∞ variation H of c fixing endpoints. First, we take an arbitrary C^∞ variation H of c with $H(u, t_i) = c(t_i)$, $i = 0, \ldots, k$, i.e., $V(t_i) = 0$. By (3.107), we obtain

$$L'(0) = -\lambda \int_a^b g_{\dot c}(\kappa, V)dt = 0, \tag{3.108}$$

which implies that the geodesic curvature vanishes,

$$\kappa(t) = 0.$$

Now for any $1 \le i_o \le k-1$ and $v \in T_{c(t_{i_o})}M$, we take a piecewise C^∞ variation H of c, that fixes two endpoints of c with

$$V(t_{i_o}) = v, \quad H(u, t_i) = c(t_i), \quad i \ne i_o. \tag{3.109}$$

By (3.107), we obtain

$$L'(0) = \frac{1}{\lambda}\sum_{i=1}^{k}\left\{g_{\dot{c}(t_{i_o}^-)}(\dot{c}(t_{i_o}^-), v) - g_{\dot{c}(t_{i_o}^+)}(\dot{c}(t_{i_o}^+), v)\right\} = 0. \tag{3.110}$$

It is easy to prove that if $g_v(y, w) = g_v(v, w)$ for $y, v, w \in T_xM$, then $y = v$. From this observation, we conclude that

$$\dot{c}(t_{i_o}^-) = \dot{c}(t_{i_o}^+).$$

That is, c is C^1 at each t_i. In local coordinates, $\kappa = 0$ is equivalent to the following system of second order differential equations:

$$\ddot{c}^i + 2G^i(\dot{c}) = 0. \tag{3.111}$$

The local functions G^i in (3.111) are called the geodesic coefficients and the vector field that they generate on $TM \setminus \{0\}$

$$G := v^i \frac{\partial}{\partial x^i} - 2G^i(x, v)\frac{\partial}{\partial v^i} \tag{3.112}$$

are called sprays induced by the metric F. We now show how these geodesic equations are similar to ones of Riemannian geometry but only differ in the Christoffel symbols. First we derive the covariant metric g_{ij} for our new homogeneous Lagrangian $\Lambda(\mathrm{x}, \mathrm{x}')$. The Finsler metric is defined as

$$g_{\mathrm{x}'^i\mathrm{x}'^j} := \frac{1}{2}\frac{\partial^2\Lambda^2}{\partial \mathrm{x}'^i \partial \mathrm{x}'^j}(\mathrm{x}, \mathrm{x}'), \qquad i, j = 0, 1, \ldots, n. \tag{3.113}$$

As usual the Lagrangian $\Lambda(\mathrm{x},\mathrm{x}')$ can be written using the metric $g_{ij}(\mathrm{x},\mathrm{x}')$ as follows:

$$\Lambda^2(\mathrm{x},\mathrm{x}') = g_{jk}(\mathrm{x},\mathrm{x}')\mathrm{x}'^j\mathrm{x}'^k. \tag{3.114}$$

Differentiating the Finsler function Λ^2 as given above, with respect to x'^i and x^i, we have

$$\begin{aligned} 2\Lambda\frac{\partial\Lambda}{\partial\mathrm{x}'^i} &= 2g_{jk}(\mathrm{x},\mathrm{x}')\frac{\partial\mathrm{x}'^j}{\partial\mathrm{x}'^i}\mathrm{x}'^k + \frac{\partial g_{jk}}{\partial\mathrm{x}'^i}(\mathrm{x},\mathrm{x}')\mathrm{x}'^j\mathrm{x}'^k, \\ &= 2g_{jk}(\mathrm{x},\mathrm{x}')\delta^j_i\mathrm{x}'^k \\ &= 2g_{ik}(\mathrm{x},\mathrm{x}')\mathrm{x}'^k \\ \frac{\partial\Lambda}{\partial\mathrm{x}'^i}(\mathrm{x},\mathrm{x}') &= \frac{1}{\Lambda}g_{ik}(\mathrm{x},\mathrm{x}')\mathrm{x}'^k. \end{aligned} \tag{3.115}$$

Similarly, by differentiating equation (3.114) with respect to x^i, we have

$$\frac{\partial\Lambda}{\partial\mathrm{x}^i}(\mathrm{x},\mathrm{x}') = \frac{1}{2\Lambda}\frac{\partial g_{jk}}{\partial\mathrm{x}^i}(\mathrm{x},\mathrm{x}')\mathrm{x}'^j\mathrm{x}'^k. \tag{3.116}$$

Substituting the above terms in the Euler-Lagrange equation, we have

$$\begin{aligned}
\frac{d}{ds}\left(\frac{\partial \Lambda}{\partial \mathrm{x}'^i}\right) - \frac{\partial \Lambda}{\partial \mathrm{x}^i} &= \frac{d}{ds}\left(\frac{1}{\Lambda} g_{ik}(\mathrm{x},\mathrm{x}')\mathrm{x}'^k\right) - \frac{1}{2\Lambda}\frac{\partial g_{jk}}{\partial \mathrm{x}^i}(\mathrm{x},\mathrm{x}')\mathrm{x}'^j \mathrm{x}'^k \\
&= \frac{1}{\Lambda}\left\{ g_{ik}\mathrm{x}''^k + \left(\frac{\partial g_{ih}}{\partial \mathrm{x}^j} - \frac{1}{2}\frac{\partial g_{jh}}{\partial \mathrm{x}^i}\right)\mathrm{x}'^j \mathrm{x}'^h \right\} - \frac{1}{\Lambda^2}\frac{d\Lambda}{ds} g_{ik}\mathrm{x}^k \\
&= \frac{1}{\Lambda}\left(g_{ik}\mathrm{x}''^k + \Gamma_{ijh}\mathrm{x}'^j \mathrm{x}'^h - \frac{1}{\Lambda}\frac{d\Lambda}{ds} g_{ik}\mathrm{x}'^k \right), \\
&= \frac{1}{\Lambda} g_{ik}\left(\mathrm{x}''^k + \Gamma^k_{jh}\mathrm{x}'^j \mathrm{x}'^h - \frac{d\log\Lambda}{ds}\mathrm{x}'^k \right), \qquad (3.117)
\end{aligned}$$

where Γ_{ijk} and Γ^i_{jk} are the usual Christoffel symbols of the first and second kind, respectively.

Since it follows from equation (3.117) that $\frac{d\Lambda}{ds} = 0$ and the condition that the inverse metric g^{ij} exists, we have

$$\frac{d^2\mathrm{x}^i}{ds^2} + \Gamma^i_{jk}\left(\mathrm{x}, \frac{d\mathrm{x}}{ds}\right)\frac{d\mathrm{x}^j}{ds}\frac{d\mathrm{x}^k}{ds} = 0, \qquad (3.118)$$

$$g_{ij}\left(\mathrm{x}, \frac{d\mathrm{x}}{ds}\right)\frac{d\mathrm{x}^i}{ds}\frac{d\mathrm{x}^j}{ds} = 1. \qquad (3.119)$$

Equation (3.118) together with constant speed condition (3.119) is called the *Finsler geodesic equation.*

Computing the inverse metric g^{ij} and the Christoffel symbols Γ^i_{jk} in the above geodesic equation is no easy task for a general problem. We now prove some theorems which facilitate in reducing the complexity of their computation.

Theorem 3.4.4. *Given the metric $g_{ij}(x, \dot{x})$ of the Finsler manifold (M, F), the determinant of the metric is given by*

$$\det(g_{ij}) = -F^{n-1} \cdot \det \begin{bmatrix} F_{\dot{x}^i \dot{x}^j} & F_{\dot{x}^i} \\ F_{\dot{x}^j} & 0 \end{bmatrix} \tag{3.120}$$

where $F_{\dot{x}^i \dot{x}^j}$ is the Hessian of the Finsler function F and n is the dimension of the manifold.

Proof. From equation (3.41), we have

$$\det(g_{ij}) = \det\left(F\, F_{\dot{x}^i \dot{x}^j} + F_{\dot{x}^i} F_{\dot{x}^j}\right) \tag{3.121}$$

$$= \det \begin{bmatrix} F\, F_{\dot{x}^i \dot{x}^j} + F_{\dot{x}^i}\, F_{\dot{x}^j} & F_{\dot{x}^i} \\ 0 & 1 \end{bmatrix} \tag{3.122}$$

$$= \det \begin{bmatrix} F\, F_{\dot{x}^i \dot{x}^j} & F_{\dot{x}^i} \\ -F_{\dot{x}^j} & 1 \end{bmatrix} \tag{3.123}$$

$$= -F^{n-1} \det \begin{bmatrix} F_{\dot{x}^i \dot{x}^j} & F_{\dot{x}^i} \\ F_{\dot{x}^j} & -F \end{bmatrix} \tag{3.124}$$

$$= -F^{n-1} \det \begin{bmatrix} F_{\dot{x}^i \dot{x}^j} & F_{\dot{x}^i} \\ F_{\dot{x}^j} & 0 \end{bmatrix} \tag{3.125}$$

In the above derivation, we subtract suitable multiples of the last column of equation (3.122) from the remaining columns to derive equation (3.123), from which we factor out F to obtain equation (3.124). Equation (3.125) is derived by applying equations (3.31) and (3.32) to (3.124). ☐

The condition that the metric g_{ij} be nonsingular implies that the matrix $F_{\dot{x}^i\dot{x}^j}$ has rank $(n-1)$. This means that the relation can be further simplified as follows.

$$\det \begin{bmatrix} F_{\dot{x}^i\dot{x}^j} & F_{\dot{x}^i} \\ F_{\dot{x}^j} & 0 \end{bmatrix} = F \cdot \det \begin{bmatrix} F_{\dot{x}^2\dot{x}^1} & \cdots & F_{\dot{x}^2\dot{x}^n} \\ \vdots & \ddots & \vdots \\ F_{\dot{x}^n\dot{x}^1} & \cdots & F_{\dot{x}^n\dot{x}^n} \\ F_{\dot{x}^1} & \cdots & F_{\dot{x}^n} \end{bmatrix} \tag{3.126}$$

$$= F^2 \cdot \det \begin{bmatrix} F_{\dot{x}^2\dot{x}^2} & \cdots & F_{\dot{x}^2\dot{x}^n} \\ \vdots & \ddots & \vdots \\ F_{\dot{x}^n\dot{x}^2} & \cdots & F_{\dot{x}^n\dot{x}^n} \end{bmatrix} \tag{3.127}$$

$$= F^2 \cdot \det H^{11} \tag{3.128}$$

Hence, the determinant reduces to

$$\det g_{ij} = -F^{n+1} \cdot \det H^{11}. \tag{3.129}$$

Theorem 3.4.5. *The inverse metric $g^{ij}(x,\dot{x})$ of the Finsler function F is given* *by*

$$g^{ij}(x,\dot{x}) = \frac{(-1)^{i+j+1}}{F^3} \frac{\det \begin{bmatrix} F_{\dot{x}^k\dot{x}^l} & F_{\dot{x}^k} \\ F_{\dot{x}^l} & -F \end{bmatrix}_{k\neq i,\, l\neq j}}{\det \begin{bmatrix} F_{\dot{x}^k\dot{x}^l} \end{bmatrix}_{k\neq i,\, l\neq j}} \tag{3.130}$$

Proof. The inverse metric g^{ij} which is the induced metric on the cotangent space T^*M, is given by

$$g^{ij} = \frac{1}{\det g_{ij}} \text{adj } g_{ij} \tag{3.131}$$

where

$$\text{adj } g_{ij} = (-1)^{i+j} \det H^{ij}$$

The elements of the adjoint matrix could be written

$$[\,\text{adj } G\,]_{ij} \;=\; (-1)^{i+j} \det \begin{bmatrix} g_{11} & \cdots & g_{1(i-1)} & g_{1(i+1)} & \cdots & g_{1n} \\ \vdots & \ddots & & & & \vdots \\ g_{(j-1)1} & & g_{(j-1)(i-1)} & g_{(j-1)(i+1)} & & \\ g_{(j+1)1} & & g_{(j+1)(i-1)} & g_{(j+1)(i+1)} & & \\ \vdots & & & & \ddots & \vdots \\ g_{n1} & \cdots & & & \cdots & g_{nn} \end{bmatrix}$$

$$= \quad (-1)^{i+j} \;\det \begin{bmatrix} F\,F_{\dot{x}^k\dot{x}^l} + F_{\dot{x}^k}F_{\dot{x}^l} & F_{\dot{x}^k} \\ 0 & 1 \end{bmatrix}_{k\neq i,\, l\neq j} \tag{3.132}$$

$$= \quad (-1)^{i+j} \;\det \begin{bmatrix} F\,F_{\dot{x}^k\dot{x}^l} & F_{\dot{x}^k} \\ -F_{\dot{x}^l} & 1 \end{bmatrix}_{k\neq i,\, l\neq j} \tag{3.133}$$

$$= \quad (-1)^{i+j+1} \cdot F^{n-2} \cdot \det \begin{bmatrix} F_{\dot{x}^k\dot{x}^l} & F_{\dot{x}^k} \\ F_{\dot{x}^l} & -F \end{bmatrix}_{k\neq i,\, l\neq j} \tag{3.134}$$

□

Theorem 3.4.6. *The geodesic coefficients $G_i(x,v)$ of the geodesic equation derived from the metric $g_{ij}(x,v)$ are given by*

$$G_i = \frac{1}{2}\left(v^j[F^2]_{x^j v^i} - [F^2]_{x^i}\right) \tag{3.135}$$

Proof. The Christoffel symbols of the first kind γ_{ijk} are given as follows

$$\gamma_{jik} = \frac{\partial g_{ij}}{\partial x^k} - \frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}$$

The geodesic coefficients can now be written as

$$G_i = \frac{\partial g_{ij}}{\partial x^k}v^j v^k - \frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}v^j v^k$$

From equation (3.38), it follows that

$$\begin{aligned}
\frac{\partial g_{ij}}{\partial x^k} v^j v^k &= \left(F F_{v^i v^j} + F_{v^i} F_{v^j}\right)_{x^k} v^j v^k \\
&= F_{x^k} F_{v^i v^j} v^j v^k + F F_{x^k v^i v^j} v^j v^k + F_{x^k v^i} F_{v^j} v^j v^k + F_{v^i} F_{x^k v^j} v^j v^k \\
&= F_{x^k v^i} F v^k + F_{v^i} F_{x^k} v^k \\
&= 2 v^k [F^2]_{x^k v^i}
\end{aligned}$$

The second term is given by equation (3.115)

$$\frac{\partial g_{jk}}{\partial x^i} v^j v^k = 2 F F_{x^i}$$

and hence we have proved the theorem. □

3.4.3 Geodesics of Optimal Control Problems

As before the Finsler geodesics are given by the following equation:

$$\frac{d^2 \mathrm{x}^i}{ds^2} + \Gamma^i_{jk}\left(\mathrm{x}, \frac{d\mathrm{x}}{ds}\right) \frac{d\mathrm{x}^j}{ds} \frac{d\mathrm{x}^k}{ds} = 0, \tag{3.136}$$

where Γ^i_{jk} are Christoffel symbol's of the second kind and $i, j, k = 1, 2, \ldots, 2n + m + 1$. The Finsler geodesics for the Optimal control problem can be written in terms of the variables (t', y', c, p) as following

Time

$$\begin{aligned}
\frac{d^2 t}{ds^2} \quad &= \quad -\Gamma^0_{00}\frac{dt}{ds}\frac{dt}{ds} - \left[\sum_{k=1}^{n}\Gamma^0_{0k}\frac{dy}{ds}^k + \sum_{k=1}^{m}\Gamma^0_{0k}c^k + \sum_{k=1}^{n}\Gamma^0_{0k}p^k\right]\frac{dt}{ds} \\
&\quad -\sum_{j=1}^{n}\left[\sum_{k=1}^{n}\Gamma^0_{jk}\frac{dy}{ds}^k + \sum_{k=1}^{m}\Gamma^0_{jk}c^k + \sum_{k=1}^{n}\Gamma^0_{jk}p^k\right]\frac{dy}{ds}^j \qquad (3.137) \\
&\quad -\sum_{j=1}^{m}\left[\sum_{k=1}^{m}\Gamma^0_{jk}c^k - \sum_{k=1}^{n}\Gamma^0_{jk}p^k\right]c^j \\
&\quad -\sum_{j=1}^{n}\sum_{j=1}^{n}\Gamma^0_{jk}p^k p^j
\end{aligned}$$

States $\qquad i = 1, \ldots, n$

$$\begin{aligned}
\frac{d^2 y^i}{ds^2} \quad &= \quad -\Gamma^i_{00}\frac{dt}{ds}\frac{dt}{ds} - \left[\sum_{k=1}^{n}\Gamma^i_{0k}\frac{dy}{ds}^k + \sum_{k=1}^{m}\Gamma^i_{0k}c^k + \sum_{k=m}^{m}\Gamma^i_{0k}p^k\right]\frac{dt}{ds} \\
&\quad -\sum_{j=1}^{n}\left[\sum_{k=1}^{n}\Gamma^i_{jk}\frac{dy}{ds}^k + \sum_{k=1}^{m}\Gamma^i_{jk}c^{k-n} + \sum_{k=1}^{n}\Gamma^i_{jk}p^k\right]\frac{dy}{ds}^j \qquad (3.138) \\
&\quad -\sum_{j=1}^{m}\left[\sum_{k=1}^{m}\Gamma^i_{jk}c^k + \sum_{k=1}^{n}\Gamma^i_{jk}p^k\right]c^j \\
&\quad -\sum_{j=1}^{n}\sum_{j=1}^{n}\Gamma^i_{jk}p^k p^j
\end{aligned}$$

Controls $\quad i = 1, \ldots, m$

$$\begin{aligned}
\frac{d\mathrm{c}^i}{ds} &= -\Gamma^i_{00}\frac{d\mathrm{t}}{ds}\frac{d\mathrm{t}}{ds} - \left[\sum_{k=1}^{n}\Gamma^i_{0k}\frac{d\mathrm{y}^k}{ds} + \sum_{k=1}^{m}\Gamma^i_{0k}\mathrm{c}^k + \sum_{k=1}^{n}\Gamma^i_{0k}\mathrm{p}^k\right]\frac{d\mathrm{t}}{ds} \\
&\quad - \sum_{j=1}^{n}\left[\sum_{k=1}^{n}\Gamma^i_{jk}\frac{d\mathrm{y}^k}{ds} + \sum_{k=1}^{m}\Gamma^i_{jk}\mathrm{c}^k + \sum_{k=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\right]\frac{d\mathrm{y}^j}{ds} \qquad (3.139) \\
&\quad - \sum_{j=1}^{m}\left[\sum_{k=1}^{m}\Gamma^i_{jk}\mathrm{c}^k - \sum_{k=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\right]\mathrm{c}^j \\
&\quad - \sum_{j=1}^{n}\sum_{j=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\mathrm{p}^j
\end{aligned}$$

Costates $\quad i = 1, \ldots, n$

$$\begin{aligned}
\frac{d\mathrm{p}^i}{ds} &= -\Gamma^i_{00}\frac{d\mathrm{t}}{ds}\frac{d\mathrm{t}}{ds} - \left[\sum_{k=1}^{n}\Gamma^i_{0k}\frac{d\mathrm{y}^k}{ds} + \sum_{k=1}^{m}\Gamma^i_{0k}\mathrm{u}^k + \sum_{k=1}^{n}\Gamma^i_{0k}\mathrm{p}^k\right]\frac{d\mathrm{t}}{ds} \\
&\quad - \sum_{j=1}^{n}\left[\sum_{k=1}^{n}\Gamma^i_{jk}\frac{d\mathrm{y}^k}{ds} + \sum_{k=1}^{m}\Gamma^i_{jk}\mathrm{c}^{k-n} + \sum_{k=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\right]\frac{d\mathrm{x}^j}{ds} \\
&\quad - \sum_{j=1}^{m}\left[\sum_{k=1}^{m}\Gamma^i_{jk}\mathrm{c}^k - \sum_{k=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\right]\mathrm{c}^{j-n} \qquad (3.140) \\
&\quad - \sum_{j=1}^{n}\sum_{j=1}^{n}\Gamma^i_{jk}\mathrm{p}^k\mathrm{p}^j
\end{aligned}$$

General Problem

Given an optimal control problem with the following data

$$\min \int_a^b l(x,u)dt$$

subject to

$$\dot{x} = f(x,u),$$

we wish to find the corresponding geodesic equations. While one could compute the geodesic equations using the metric g_{ij} it is a computationally inefficient technique when one deals with problems of large dimension in states, controls, and constraints, we instead make use of the above derived formulas for the metric inverse g^{ij} and the geodesic spray coefficients G_i. We begin with constructing the Lagrangian Λ for the above optimal control problem in the augmented state tangent space $(\mathrm{x}, \mathrm{x}')$ as follows

$$\Lambda(\mathrm{x},\mathrm{x}') := \mathrm{t}' \cdot l\left(\mathrm{y}, \frac{\mathrm{c}}{\mathrm{t}'}\right) + \frac{1}{\mathrm{t}'}\left\langle \mathrm{p}^\top \, , \, \mathrm{y}' - \mathrm{t}' \cdot f\left(\mathrm{y}, \frac{\mathrm{c}}{\mathrm{t}'}\right)\right\rangle . \tag{3.141}$$

We now compute some useful derivatives for the computation of metric inverse and spray coefficients which lead to the geodesic equation.

Theorem 3.4.7. *Given an optimal control problem with metric (3.141), the dual variables are given by*

$$^{y}z_i = \frac{\Lambda}{t'}p^i, \quad {}^{c}z_i = \frac{\Lambda}{t'}H_{c^i}, \quad {}^{p}z_i = \frac{\Lambda}{t'}(y'^i - t'f^i), \tag{3.142}$$

Proof.

$$\Lambda_{\mathrm{x}'^i} = \begin{bmatrix} \Lambda_{\mathrm{t}'} \\ \Lambda_{\mathrm{y}'^i} \\ \Lambda_{\mathrm{c}^i} \\ \Lambda_{\mathrm{p}^i} \end{bmatrix} = \begin{bmatrix} l - \frac{1}{\mathrm{t}'}\mathrm{p}^j\mathrm{y}'^j - \frac{\mathrm{c}}{\mathrm{t}'^2}(\mathrm{t}' l_{\mathrm{t}'} - \mathrm{p}^j f^j_{\mathrm{t}'}) \\ \frac{1}{\mathrm{t}'}\mathrm{p}^i \\ \frac{1}{\mathrm{t}'}(\mathrm{t}' l_{\mathrm{c}^i} - \mathrm{p}^j f^j_{\mathrm{c}^i}) \\ \frac{1}{\mathrm{t}'}(\mathrm{y}'^i - \mathrm{t}' f^i) \end{bmatrix} \tag{3.143}$$

Since the dual variables are given by

$$z_i = g_{ij} x'^j = \Lambda \frac{\partial \Lambda}{\partial \mathrm{x}'^i} \tag{3.144}$$

we have

$$z_i = \Lambda \begin{bmatrix} l - \frac{1}{\mathrm{t}'}\mathrm{p}^j\mathrm{y}'^j - \frac{\mathrm{c}}{\mathrm{t}'} H_{t'} \\ H_{x'^i} \\ H_{c^i} \\ H_{p^i} \end{bmatrix} \tag{3.145}$$

□

The Hessian of the Lagrangian Λ in the tangent space x' is given by

$$\frac{\partial^2 \Lambda}{\partial \mathrm{x}' \partial \mathrm{x}'} = \begin{bmatrix} 2\frac{\mathrm{p}}{\mathrm{t}'^3}(\mathrm{y}' - \mathrm{c} f_{\mathrm{t}'}) + \frac{\mathrm{c}^2}{\mathrm{t}'^4}(\mathrm{t}' l_{\mathrm{t}'\mathrm{t}'} - \mathrm{p} f_{\mathrm{t}'\mathrm{t}'}) & -\frac{1}{\mathrm{t}'}[\mathrm{p}] & [l_{\mathrm{t}'\mathrm{c}^i} - \mathrm{p} f_{\mathrm{t}'\mathrm{c}^i}] & -\frac{1}{\mathrm{t}'}[\mathrm{y}' - \mathrm{t}' \cdot f_{\mathrm{t}'}] \\ -\frac{1}{\mathrm{t}'}[\mathrm{p}] & [\mathbf{0}]_{n\times n} & [\mathbf{0}]_{n\times m} & [I]_{n\times n} \\ [l_{\mathrm{t}'\mathrm{c}^i} - \mathrm{p} f_{\mathrm{t}'\mathrm{c}^i}] & [\mathbf{0}]_{m\times n} & [t' l_{\mathrm{c}^i\mathrm{c}^j} - p f_{\mathrm{c}^i\mathrm{c}^j}]_{m\times m} & -[f_{\mathrm{c}^i}]_{m\times n} \\ -\frac{1}{\mathrm{t}'}[\mathrm{y}' - \mathrm{t}' \cdot f_{\mathrm{t}'}] & [I]_{n\times n} & -[f_{\mathrm{c}^i}]_{n\times m} & [\mathbf{0}]_{n\times n} \end{bmatrix} \tag{3.146}$$

Theorem 3.4.8. *Given the optimal control problem the metric g_{ij} is invertible iff*

$$\det H_{c^i c^j} \neq 0, \tag{3.147}$$

where

$$H_{c^i c^j} \;=\; [\, t' l_{\mathrm{c}^i \mathrm{c}^j} - p f_{\mathrm{c}^i \mathrm{c}^j} \,]\,. \tag{3.148}$$

Proof. We use the above computed Hessian of the Lagrangian in the tangent space x' to compute the determinant of the metric g_{ij} as follows

$$\det g_{ij} \;=\; -\frac{\Lambda^{n+1}}{\mathrm{t}'^{n}} \det \begin{bmatrix} [\mathbf{0}]_{n\times n} & [\mathbf{0}]_{n\times m} & [I]_{n\times n} \\ [\mathbf{0}]_{m\times n} & [\mathrm{t}' l_{\mathrm{c}^i\mathrm{c}^j} - \mathrm{p} f_{\mathrm{c}^i\mathrm{c}^j}]_{m\times m} & -[f_{\mathrm{c}^i}]_{m\times n} \\ [I]_{n\times n} & -[f_{\mathrm{c}^i}]_{n\times m} & [\mathbf{0}]_{n\times n} \end{bmatrix} \tag{3.149}$$

$$\;=\; -\frac{\Lambda^{n+1}}{\mathrm{t}'^{n}} \det [\mathrm{t}' l_{\mathrm{c}^i\mathrm{c}^j} - \mathrm{p} f_{\mathrm{c}^i\mathrm{c}^j}]_{m\times m} \tag{3.150}$$

From the determinant of the metric g_{ij}, it is quite evident that

$$\det g_{ij} \;\neq\; 0 \;\Leftrightarrow\; \det [\, \mathrm{t}' l_{\mathrm{c}^i\mathrm{c}^j} - \mathrm{p} f_{\mathrm{c}^i\mathrm{c}^j} \,] \;\neq\; 0. \tag{3.151}$$

Hence, we require that the matrix $\mathrm{t}' l_{\mathrm{c}^i\mathrm{c}^j} - \mathrm{p} f_{\mathrm{c}^i\mathrm{c}^j}$ be nonsingular. ☐

From the following matrices we compute the geodesic coefficients G_i of the optimal control problem with the metric Λ as follows:

$$\Lambda_{\mathrm{x}^i} = \begin{bmatrix} 0 \\ \mathrm{t}' l_{\mathrm{y}^i} - \mathrm{p} f_{\mathrm{y}^i} \\ 0 \\ 0 \end{bmatrix} \tag{3.152}$$

$$\Lambda_{\mathrm{x}'^i\mathrm{x}^k} = \begin{bmatrix} 0 & l_{\mathrm{y}^i} - \frac{\mathrm{c}}{\mathrm{t}'^2}(\mathrm{t}' l_{\mathrm{y}^i\mathrm{t}'} - \mathrm{p} f_{\mathrm{y}^i\mathrm{t}'}) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & l_{\mathrm{y}^i\mathrm{c}^k} - \mathrm{p}^j f^j_{\mathrm{y}^i\mathrm{c}^k} & 0 & 0 \\ 0 & -f^k_{\mathrm{y}^i} & 0 & 0 \end{bmatrix} \tag{3.153}$$

We then use Theorem 3.4.6 and equations (3.152) and (3.153), we have

$$\begin{aligned}
{}^{\mathrm{t}}G_0 &= \Lambda\Lambda_{\mathrm{t}'\mathrm{y}^k}\mathrm{y}'^k + \Lambda_{\mathrm{t}'}\Lambda_{\mathrm{y}^k}\mathrm{y}'^k \\
&= \Lambda(l_{\mathrm{y}^k} - \frac{c}{\mathrm{t}'^2}(\mathrm{t}' l_{\mathrm{y}^k t'} - \mathrm{p} f_{\mathrm{y}^k\mathrm{t}'}))\mathrm{y}'^k + (l - \frac{1}{\mathrm{t}'}\mathrm{p}^j \mathrm{y}'^j - \frac{\mathrm{c}}{\mathrm{t}'^2}(\mathrm{t}' l_{\mathrm{t}'} - \mathrm{p}^j f^j_{\mathrm{t}'}))(l_{\mathrm{y}^k} - \mathrm{p}^j f^j_{\mathrm{y}^k})\mathrm{y}'^k \\
{}^{\mathrm{y}}G_i &= \Lambda_{\mathrm{y}'^i}\Lambda_{\mathrm{y}^k}\mathrm{y}'^k - \Lambda\Lambda_{\mathrm{y}^i} \\
&= \frac{\mathrm{p}^i}{\mathrm{t}'}(l_{\mathrm{y}^k} - \mathrm{p}^j f^j_{\mathrm{y}^k})\mathrm{y}'^k - \Lambda(l_{\mathrm{y}^i} - \mathrm{p}^j f^j_{\mathrm{y}^i}) \\
{}^{\mathrm{c}}G_i &= \Lambda\Lambda_{\mathrm{c}^i\mathrm{y}^k}\mathrm{y}'^k + \Lambda_{\mathrm{c}^i}\Lambda_{\mathrm{y}^k}\mathrm{y}'^k \\
&= \Lambda(l_{\mathrm{y}^k\mathrm{c}^i} - \mathrm{p}^j f^j_{\mathrm{y}^k\mathrm{c}^i})\mathrm{y}'^k + (l_{\mathrm{c}^i} - \mathrm{p}^k f^k_{\mathrm{c}^i})(l_{\mathrm{y}^k} - \mathrm{p}^j f^j_{\mathrm{y}^k})\mathrm{y}'^k \\
{}^{\mathrm{p}}G_i &= \Lambda\Lambda_{\mathrm{p}^i\mathrm{y}^k}\mathrm{y}'^k + \Lambda_{\mathrm{p}^i}\Lambda_{\mathrm{y}^k}\mathrm{y}'^k \\
&= -\Lambda f^i_{\mathrm{y}^k}\mathrm{y}'^k + (\mathrm{y}'^i - \mathrm{t}' f^i)(l_{\mathrm{y}^k} - \mathrm{p}^j f^j_{\mathrm{y}^k})\mathrm{y}'^k
\end{aligned}$$

Example 6. *Linear-Quadratic Problem:The Linear Quadratic Optimal control can be described as follows:*

$$\min_u \int_0^T x^\top Rx + u^\top Pu$$

given the linear control system

$$\dot{x} = Ax + Bu, \quad x(0) = x_0 \text{ and } x(T) = x_T.$$

Here the matrices P and R are positive-definite. The augmented Lagrangian for the LQR problem is then given by

$$\Lambda(\mathrm{x}, \mathrm{x}') := \frac{1}{2\mathrm{t}'} \left[\mathrm{c}^\top P\mathrm{c} + \mathrm{t}'^2\, \mathrm{y}^\top R\mathrm{y}\right] + \frac{1}{\mathrm{t}'} \left\langle \mathrm{p}^\top \,,\, \mathrm{y}' - \mathrm{t}' A\mathrm{y} - B\mathrm{c}\right\rangle \tag{3.154}$$

Computing the Hessian

$$\frac{\partial^2 \Lambda}{\partial \mathrm{x}' \partial \mathrm{x}'} = \frac{1}{\mathrm{t}'} \begin{bmatrix} \frac{1}{\mathrm{t}'^2}(\mathrm{c}^\top P\mathrm{c} + \mathrm{p}(\mathrm{y}' - B\mathrm{c})) & -\frac{1}{\mathrm{t}'} & -\frac{1}{\mathrm{t}'}(\mathrm{c}^\top P + P\mathrm{c} - \mathrm{p}B) & -\frac{1}{\mathrm{t}'}(\mathrm{y}' - B\mathrm{c}) \\ -\frac{1}{\mathrm{t}'} & [\mathbf{0}]_{n\times n} & [\mathbf{0}]_{n\times m} & [I]_{n\times n} \\ -\frac{1}{\mathrm{t}'}(\mathrm{c}^\top P + P\mathrm{c} - \mathrm{p}B) & [\mathbf{0}]_{m\times n} & [P]_{m\times m} & -[B]_{m\times n} \\ -\frac{1}{\mathrm{t}'}(\mathrm{y}' - B\mathrm{c}) & [I]_{n\times n} & -[B]_{n\times m} & [\mathbf{0}]_{n\times n} \end{bmatrix} \tag{3.155}$$

and the determinant of the metric

$$\det g_{ij} = -\frac{\Lambda^{n+1}}{\mathrm{t}'^n} \det \begin{bmatrix} [\mathbf{0}]_{n\times n} & [\mathbf{0}]_{n\times m} & [I]_{n\times n} \\ [\mathbf{0}]_{m\times n} & [P]_{m\times m} & -[B]_{m\times n} \\ [I]_{n\times n} & -[B]_{n\times m} & [\mathbf{0}]_{n\times n} \end{bmatrix} \tag{3.156}$$

we can conclude that

$$\det g_{ij} \;\neq\; 0 \;\Leftrightarrow\; \det P \;\neq\; 0. \tag{3.157}$$

Since the matrix P be is positive definite we conclude that the metric g_{ij} is invertible.

We first compute the following matrices to

$$\frac{\partial\Lambda}{\partial \mathrm{x}'} = \begin{bmatrix} \frac{1}{\mathrm{t}'^2}(\mathrm{c}^\top P\mathrm{c} + \mathrm{p}(\mathrm{y}' - B\mathrm{c})) \\ \frac{1}{\mathrm{t}'}\mathrm{p} \\ -\frac{1}{\mathrm{t}'}(\mathrm{c}^\top P + P\mathrm{c} - \mathrm{p}B) \\ -\frac{1}{\mathrm{t}'}(\mathrm{y}' - \mathrm{t}'A\mathrm{y} - B\mathrm{c}) \end{bmatrix} \tag{3.158}$$

$$\frac{\partial\Lambda}{\partial \mathrm{x}} = \begin{bmatrix} 0 \\ \frac{\mathrm{t}'}{4}(\mathrm{y}^\top R + R\mathrm{y}) - \mathrm{p}^\top A \\ 0 \\ 0 \end{bmatrix} \tag{3.159}$$

to get the geodesic coefficients for the Linear Quadratic Optimal Control Problem which are given by

$$^{\mathrm{t}}G_0 = \Lambda(l_{y^k} - \frac{c}{t'^2}(t'l_{y^k t'} - pf_{y^k t'}))y'^k + (l - \frac{1}{t'}p^j y'^j - \frac{c}{t'^2}(t'l_{t'} - p^j f^j_{t'}))(l_{y^k} - p^j f^j_{y^k})y'^k$$

$$^{\mathrm{y}}G = \frac{1}{t'}\mathrm{p}(\mathrm{y}^\top R + R\mathrm{y} - \mathrm{p}^\top A)\mathrm{y}' - \Lambda(\mathrm{y}^\top R + R\mathrm{y} - \mathrm{p}^\top A)$$

$$^{\mathrm{c}}G = (\mathrm{c}^\top P + P\mathrm{c} - \mathrm{p}B)(\mathrm{y}^\top R + R\mathrm{y} - \mathrm{p}^\top A)\mathrm{y}'$$

$$^{\mathrm{p}}G = -\Lambda A\mathrm{y}' + (\mathrm{y}' - t'A\mathrm{y} - B\mathrm{c})(R\mathrm{y} - \mathrm{p}A)\mathrm{y}'$$

Example 7. *Given the Linear Quadratic optimal control problem*

$$l = \frac{1}{2}u^2 + x^2,$$

$$\dot{x} = x + u$$

we wish to compute its geodesic equations. Notice that the above problem could be written as a Calculus of Variations problem with the following Lagrangian

$$\begin{aligned} L &= \frac{1}{2}(\dot{x} - x)^2 + x^2 \\ &= \frac{1}{2}(\dot{x}^2 - 2x\dot{x} + 3x^2) \end{aligned}$$

Its quite evident that the Euler-Lagrange equation for the above Calculus of Variations problem is given by

$$\ddot{x} = 3x.$$

If however, we were to derive the Finsler geodesics for the above Calculus of Variations problem then its clear that the Lagrangian is not positive homogeneous in $\dot{x}$ of degree 1. Hence we homogenize L in the following way:

$$\Lambda(t, \mathrm{x}, t', \mathrm{x}') = \frac{1}{2}\left(\frac{\mathrm{x}'^2}{t'^2} - 2\mathrm{x}\frac{\mathrm{x}'}{t'} + 3\mathrm{x}^2\right) \cdot t'$$

From which the metric is derived first

$$G(t, \mathrm{x}, t', \mathrm{x}') = \begin{bmatrix} 1/4\,\frac{3\,\mathrm{x}'^4 - 4\,\mathrm{x}'^3\mathrm{x}\,t' + 9\,\mathrm{x}^4 t'^4}{t'^4} & -1/2\,\frac{2\,\mathrm{x}'^3 - 3\,\mathrm{x}\,\mathrm{x}'^2 t' + 3\,\mathrm{x}^3 t'^3}{t'^3} \\ -1/2\,\frac{2\,\mathrm{x}'^3 - 3\,\mathrm{x}\,\mathrm{x}'^2 t' + 3\,\mathrm{x}^3 t'^3}{t'^3} & 1/2\,\frac{3\,\mathrm{x}'^2 - 6\,\mathrm{x}\,\mathrm{x}'\,t' + 5\,\mathrm{x}^2 t'^2}{t'^2} \end{bmatrix}$$

The geodesic equation is then derived

$$t'' = 2\,\frac{\left(\mathrm{x}'^2 - 6\,\mathrm{x}\,\mathrm{x}'\,t' + 3\,\mathrm{x}^2 t'^2\right) t'^2}{\mathrm{x}'^2 - 2\,\mathrm{x}\,\mathrm{x}'\,t' + 3\,\mathrm{x}^2 t'^2}$$

$$\mathrm{x}'' = \frac{\left(2\,\mathrm{x}'^3 - 9\,\mathrm{x}\,\mathrm{x}'^2 t' + 9\,\mathrm{x}^3 t'^3\right) t'}{\mathrm{x}'^2 - 2\,\mathrm{x}'\,\mathrm{x}'\,t' + 3\,\mathrm{x}^2 t'^2}$$

Let us now derive the controls for the original optimal control problem. The homogenized augmented Lagrangian is given as

$$\left(\frac{1}{2}\,\frac{\mathrm{u}^2}{\mathrm{t}'^2} + \mathrm{x}'^2 + \frac{\mathrm{p}}{\mathrm{t}'}\left(\frac{\mathrm{x}'}{\mathrm{t}'} - \mathrm{x} - \frac{\mathrm{u}}{\mathrm{t}'}\right)\right)\mathrm{t}'. \tag{3.160}$$

From the above Lagrangian we derive the geodesic equations as follows:

$$\mathrm{t}'' = 2\,\frac{\left(\mathrm{p}\,\mathrm{x}' + 2\,\mathrm{x}^2\mathrm{t}'^2 - 4\,\mathrm{x}'\,\mathrm{x}\,\mathrm{t}' - \mathrm{p}\,\mathrm{x}\,\mathrm{t}'\right)\mathrm{t}'^2}{\mathrm{u}^2 - 2\,\mathrm{u}\,\mathrm{p} - 2\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}' + 2\,\mathrm{p}\,\mathrm{x}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

$$\mathrm{x}'' = \frac{\left(2\,\mathrm{u}^2\mathrm{x}\,\mathrm{t}' - \mathrm{u}^2\mathrm{p} + \mathrm{u}^2\mathrm{x}' - 4\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}'\,\mathrm{u} - 2\,\mathrm{p}\,\mathrm{x}'\,\mathrm{u} + 2\,\mathrm{u}\,\mathrm{p}^2 - 8\,\mathrm{x}'^2\mathrm{x}\,\mathrm{t}'\right)\mathrm{t}'}{\mathrm{u}^2 - 2\,\mathrm{u}\,\mathrm{p} - 2\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}' + 2\,\mathrm{p}\,\mathrm{x}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

$$+\frac{\left(2\,\mathrm{p}^2\mathrm{x}\,\mathrm{t}' - 6\,\mathrm{p}\,\mathrm{x}^2\mathrm{t}'^2 + 4\,\mathrm{p}\,\mathrm{x}'^2 + 6\,\mathrm{x}^2\mathrm{t}'^2\mathrm{x}' - 2\,\mathrm{p}^2\mathrm{x}' + 4\,\mathrm{x}^3\mathrm{t}'^3\right)\mathrm{t}'}{\mathrm{u}^2 - 2\,\mathrm{u}\,\mathrm{p} - 2\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}' + 2\,\mathrm{p}\,\mathrm{x}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

$$\mathrm{u}' = \frac{\left(2\,\mathrm{u}^2\mathrm{xt}' - \mathrm{u}^2\mathrm{p} + 2\,\mathrm{px}'\,\mathrm{u} - 8\,\mathrm{ux}'\,\mathrm{xt}' + 4\,\mathrm{x}^2\mathrm{t}'^2\mathrm{u} + 2\,\mathrm{p}^2\mathrm{u} - 6\,\mathrm{upxt}'\right)\mathrm{t}'}{\mathrm{u}^2 - 2\,\mathrm{up} + 2\,\mathrm{px}' - 2\,\mathrm{pxt}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

$$-\frac{\left(2\,\mathrm{p}^2\mathrm{x}' + 4\,\mathrm{px}'\,\mathrm{xt}' + 4\,\mathrm{x}^3\mathrm{t}'^3 + 2\,\mathrm{p}^2\mathrm{xt}' - 6\,\mathrm{px}^2\mathrm{t}'^2\right)\mathrm{t}'}{\mathrm{u}^2 - 2\,\mathrm{up} + 2\,\mathrm{px}' - 2\,\mathrm{pxt}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

$$\mathrm{p}' = \frac{\left(2\,\mathrm{u}^2\mathrm{xt}' - \mathrm{u}^2\mathrm{p} - 4\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}'\,\mathrm{u} + 2\,\mathrm{p}^2\mathrm{u} - 2\,\mathrm{p}\,\mathrm{x}^2\mathrm{t}'^2 - 4\,\mathrm{p}\,\mathrm{x}'\,\mathrm{x}\,\mathrm{t}' + 4\,\mathrm{x}^3\mathrm{t}'^3\right)\mathrm{t}'}{\mathrm{u}^2 - 2\,\mathrm{u}\,\mathrm{p} - 2\,\mathrm{p}\,\mathrm{x}\,\mathrm{t}' + 2\,\mathrm{p}\,\mathrm{x}' + 2\,\mathrm{x}^2\mathrm{t}'^2}$$

In Figure 3.7 we have shown the optimal trajectories of the state y(s) and the

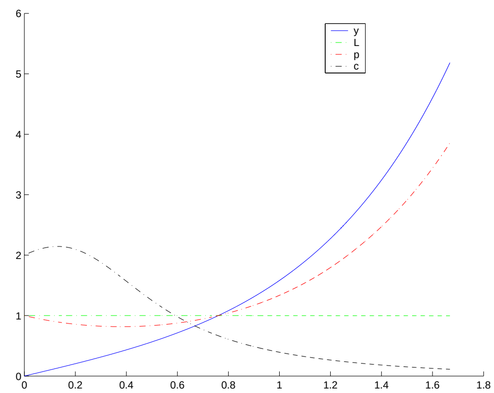

Figure 3.7: Optimal Trajectories of Example 7

controls c(s) and adjoint trajectory p(s) with boundary conditions $x_0 = 0$ and $x_{t_f} = 5.3$. We also show the trajectory for the Lagrangian L which remains constant throughout the interval.

3.5 Constrained Geodesics

So far we have computed the geodesics for optimal control problem with the control space being an open set, however in many problems of interest the control space is closed and bounded. In this section we obtain the geodesic equations under constraints. As usual we shall deal with both state and control equality and inequality constraints.

Before we do any computations we need to look carefully as to how to homogenize our augmented Lagrangians in the presence of constraints. The augmented Lagrangian in the presence of an equality constraint is written as follows:

$$\hat{L}\left(t, x, \frac{dx}{dt}\right) = L\left(t, x, \frac{dx}{dt}\right) + \lambda_i F^i\left(t, x, \frac{dx}{dt}\right) + \mu_i c^i\left(t, x, \frac{dx}{dt}\right)$$

As before, we shall now homogenize the Lagrangian in $\dot{x}$ of degree one, and the homogenized Lagrangian $\tilde{\Lambda}$ is given by

$$\begin{aligned} \tilde{\Lambda} &= \Lambda(\mathrm{x}, \mathrm{x}') + \mu_i c^i\left(t, x, \frac{dx}{ds} \cdot \left(\frac{dt}{ds}\right)^{-1}\right)\frac{dt}{ds}, \\ &= \Lambda(\mathrm{x}, \mathrm{x}') + \mu_i(s) C^i\left(\mathrm{x}, \mathrm{x}'\right), \end{aligned}$$

where

$$C^i(\mathrm{x}, \mathrm{x}') = c^i\left(t, x, \frac{dx}{ds} \cdot \left(\frac{dt}{ds}\right)^{-1}\right)\frac{dt}{ds}.$$

The problem can now be stated as

$$s = \min \int_{t_0}^{t_1} \left(g_{ij}(\mathrm{x}, \mathrm{x}')\frac{d\mathrm{x}^i}{ds}\frac{d\mathrm{x}^j}{ds}\right)^{\frac{1}{2}} ds, \tag{3.161}$$

subject to

$$C^k(\mathrm{x}, \mathrm{x}') = 0, \qquad \text{for } k = 1, \ldots, r < n. \tag{3.162}$$

3.5.1 State Equality Constraints

We now construct the Lagrangian for the above optimization problem with Lagrange multipliers μ_i, for $i = 1, \ldots, r_1$, as following

$$L(\mathrm{x}, \mathrm{x}') = (g_{ij}\mathrm{x}'^i\mathrm{x}'^j)^{\frac{1}{2}} + \sum_{i=1}^{r_1} \mu_i H^i(\mathrm{x}). \tag{3.163}$$

From the Euler-Lagrange equation for the above Lagrangian we obtain

$$g_{ik}\mathrm{x}^k_{ss} - \frac{1}{2}\left[\frac{\partial g_{kl}}{\partial \mathrm{x}^i} - \frac{\partial g_{ik}}{\partial \mathrm{x}^l} - \frac{\partial g_{il}}{\partial \mathrm{x}^k}\right]\mathrm{x}^k_s\mathrm{x}^l_s = \mu_p H^p_i, \tag{3.164}$$

where

$$H^p_i = \frac{\partial H^p}{\partial \mathrm{x}^i}.$$

Substituting the Christoffel symbols into equation (3.164), we have

$$\mathrm{x}^i_{ss} + \Gamma^i_{kl}\mathrm{x}^k_s\mathrm{x}^l_s = \mu_q g^{ik} H^q_k. \tag{3.165}$$

However the Lagrange multipliers are still unknown in the above constrained geodesic. We now compute μ_i by the requirement that the second derivatives of the constraint function with respect to arc-length vanish along the geodesic, i.e.,

$$\begin{aligned} H^p_{ss} &= H^p_{kl}\mathrm{x}^k_s\mathrm{x}^l_s + H^p_k\mathrm{x}^k_{ss} \\ &= 0. \end{aligned} \tag{3.166}$$

We now eliminate x^i_{ss} from the equation (3.165) and compute μ_i as follows

$$\begin{aligned}(g^{kl}H^p_kH^q_l)\mu_q &= H^p_i\mathrm{x}^i_{ss} + \Gamma^m_{kl}H^p_m\mathrm{x}^k_s\mathrm{x}^l_s, \\ &= -H^p_{kl}\mathrm{x}^k_s\mathrm{x}^l_s + \Gamma^m_{kl}H^p_m\mathrm{x}^k_s\mathrm{x}^l_s, \qquad (3.167)\end{aligned}$$

We now introduce the symmetric matrix h_{pq} as

$$h_{pq} = g^{kl}H^p_kH^q_l \qquad (3.168)$$

which represents the inner product of the normals to the constraint surfaces and define the contravariant form h_{pq} as

$$h^{pr}h_{rq} = \delta^p_q. \qquad (3.169)$$

From equation (3.167) we obtain the explicit expression for the Lagrange multipliers

$$\mu_p = -h^{pq}(H^p_{kl}\mathrm{x}^k_s\mathrm{x}^l_s - \Gamma^m_{kl}H^p_m\mathrm{x}^k_s\mathrm{x}^l_s). \qquad (3.170)$$

Substituting the Lagrange multipliers back into constrained geodesic equation (3.165), we get the geodesic equation free of the Lagrange multipliers, i.e.,

$$\mathrm{x}^i_{ss} + \Gamma^i_{kl}\mathrm{x}^k_s\mathrm{x}^l_s = -h^{pq}\alpha_p\beta_q \qquad (3.171)$$

where

$$\begin{aligned}\alpha_p &= g^{ik}H^p_k, \qquad (3.172)\\ \beta_p &= (H^p_{kl} - \Gamma^m_{kl}H^p_m)\mathrm{x}^k_s\mathrm{x}^l_s. \qquad (3.173)\end{aligned}$$

We further note that the constrained geodesic equation (3.171) can be written as

$$\mathrm{x}^i_{ss} + \tilde{\Gamma}^i_{kl}\mathrm{x}^k_s\mathrm{x}^l_s \quad = \quad 0, \tag{3.174}$$

where

$$\tilde{\Gamma}^i_{kl} = \Gamma^i_{kl} + h^{pq}g^{ik}H^p_k(H^p_{kl} - \Gamma^m_{kl}H^p_m).$$

3.5.2 State-Control Inequality Constraints

In the case of optimal control problems subjected to inequality constraints of the form

$$C^i(t, x, u) \leq 0, \qquad\qquad i = 1, \ldots, r,$$

one usually introduces a slack variable α and convert the inequality constraint into an equality constraint, i.e.,

$$C(t, x, u, \alpha) \equiv C(t, x, u) + \alpha^2 = 0.$$

Note that the slack variables now become controls and hence the dimension of the control space is now $m + r$, where r is the number of inequality constraints. We can then form a new Lagrangian with the control variables as $v = (u, \alpha)$. Some simple forms of these type of constraints are

$$u^i \geq u^i_{min} \qquad\qquad i = 1, \ldots m,$$

or

$$u^i_{min} \leq u^i \leq u^i_{max}, \qquad\qquad i = 1, \ldots, m.$$

In these special cases, one can introduce new unbounded controls in the following way

$$u = u_{min} + \alpha^2$$

and

$$u = u_{min} + (u_{max} - u_{min}) \sin^2(\alpha).$$

As the number of inequality constraints increases the computation of geodesics in this formalism becomes increasingly complex. So it is probably worthwhile to partition the state space and construct a possibly singular local Finsler metric on the tangent bundle. This technique of constructing metrics on the tangent space in the presence of inequality constraints was considered in [49]. We would like to continue on their work and construct local metrics instead of global metrics. This kind of constructing local Euler-Lagrange equations is also done in problems of calculus of variations when the Lagrangian is nonconvex in the velocity variable.

Chapter 4

Parallel Transport of Controls

4.1 Introduction

Connections, covariant derivatives, parallel transport, are equivalent notions which were introduced into the geometry of manifolds by Ricci and his student Levi Civita in about 1900 to give meaning to the notion that a tangent line at one point on a curve is parallel to a tangent line at another point on that same curve. There can be many connections on a manifold. The notion cannot be defined in terms of intrinsic manifold structure alone or in terms of a metric ground form such as in Riemannian geometry, where the metric ground form is positive definite, or in terms of the metric ground form $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 - dt^2$ of Einstein's four dimensional space-time. Einstein used a connection defined by a "covariant differentiation" to transfer notions of length and angle from one point to another along space time trajectories. In the 1970's, connections became ubiquitous in physics as the proper expression of the notion of physical field. The current abstract coordinate-free definition was given by Kozul in the 1950's(see [51]).

The oldest example, from Levi Civita, is the Euclidean connection on an open subset U in Euclidean space $\mathbb{R}^n$, is defined as a "covariant derivative"

$$D_uV\Big|_x = \left(u^i\frac{\partial V^1}{\partial x^i},\ldots,u^i\frac{\partial V^n}{\partial x^i}\right),$$

where $V = (V^1,\ldots,V^n) : \mathcal{U} \to \mathbb{R}^n$ is a C^∞ vector field on $\mathcal{U}$ and $u = (u^1,\ldots,u^n) \in T_x\mathcal{U} = \mathbb{R}^n$ is a tangent vector at $x \in \mathcal{U}$.

The Euclidean connection for a manifold embedded in Euclidean space may be thought of as dragging a tangent vector parallel to itself along the curve. But this description is dependent on the embedding of the manifold in Euclidean space. To get an intrinsic definition, Levi Civita looked at parallel transport by the connection of the coordinate lines of a Euclidean coordinate system at a point on a curve to an infinitisimally close point of the curve. It is an infinitisimal rotation. Integrating the equations for that rotation gave the equations of parallel transport. They describe how Euclidean coordinates change as one moves along a curve by a semigroup of invertible linear transformations along the curve. The covariant derivative can be so described.

In studying the change of the vector field $V \in C^\infty(TM)$ along a given curve C we have to consider two factors as we pass from p to q, where $p, q \in M$. First,:

the change $dV^i = \frac{dV^i}{dt}\, dt$ in $V^i(t)$ which depends solely on the definition of the vector field V^i and is independent of the metric of the space. Second, the change in metric from tangent space T_pM to tangent space T_qM. In the optimal control, the variational equation and the adjoint equation of the Pontryagin Maximum Principle are parallel transport equations for vectors and covectors, respectively. In this chapter we derive the Finsler versions which allow us to transport controls along state trajectories.

Unlike the case of manifolds given embedded in Euclidean space, where the embedding gives a preferred affine connection, the Levi-Civita or Euclidean connection, there is no such preferred choice in Finsler manifolds. Many connections have been introduced in the literature. It was Ludwig Berwald who first introduced the notion of connection for Finsler metrics. The best known connections are the Cartan connection (see [26],[64]), the Chern-Rund connection (see [8]), and the Berwald connection (see [13], [65]). The last mentioned connection is defined on sprays which are integral curves of second order ordinary differential equations(SODES) which have homogeneous right hand sides of degree 2. The first two connections are metric based connections. The Cartan connection is metric compatible and not torsion free. The Chern-Rund connection is torsion free but almost metric compatible. That each optimal control problem gives rise to a corresponding notion of geodesic and connection in Finsler Geometry was noted by Kohn-Nerode in the volume Hybrid Systems II (1994) and exploited in later papers and applications. In the Pontryagin formulation local connections were studied by Sussmann(see [67], [68]). Sussmann has used Lie brackets as connections along curves. Parallel transport using these connections then give the usual variational and adjoint equations. In this chapter we deal with Chern-Rund connection for optimal control problems.

4.2 Sprays

We notice that Finsler geodesics are given as solutions of second order differential equations(SODEs). These are in particular a homogeneous system of SODEs which are known in the differential geometry literature as sprays. One can study such SODEs which are more general than the geodesic equations obtained from Finsler metrics. We first define a spray as follows:

Definition 4.2.1. *Let M be a manifold. A spray on M is a smooth vector field G on $TM \setminus \{0\}$ expressed in a standard local coordinate system (x^i, v^i) in TM as follows*

$$\mathbb{G} = v^i \frac{\partial}{\partial x^i} - 2G^i(x,v)\frac{\partial}{\partial v^i}, \tag{4.1}$$

where $G^i(x,v)$ are local functions on TM satisfying

$$G^i(x,\lambda v) = \lambda^2 G^i(x,v), \qquad \lambda > 0. \tag{4.2}$$

A manifold with a spray is called a spray space.

Definition 4.2.2. *A regular curve c in M is called a geodesic of $\mathbb{G}$ if it is the projection of an integral curve of $\mathbb{G}$. We shall call G^i the spray coefficients of $\mathbb{G}$.*

A spray G on a manifold M determines a collection of parameterized curves (geodesics) in M. Conversely, a collection of parameterized curves with certain properties induces a spray. Let $\mathcal{G}$ be a collection of C^∞ parameterized curves $\sigma : (a, b) \to M$ with the following properties:

1. For every vector $v \in T_xM$ and any t_0, there is a curve $\gamma : (a, b) \to M$ in $\mathcal{G}$ with $t_0 \in (a, b)$ and $\dot{\gamma}(t_0) = v$;

2. For any two curves $\gamma(t)$, $a < t < b$ and $\sigma(t)$, $c < t < d$, if at some $t_0 \in (a, b)$ and $t_1 \in (c, d)$, $\dot{\gamma}(t_0) = \dot{\sigma}(t_1)$, then

$$\gamma(t_0 + t) = \sigma(t_1 + t), \qquad t \in (a - t_0, b - t_0) \cap (c - t_1, d - t_1).$$

3. For any curve $\gamma : (a, b) \to M$ in G and any $t_0 \in R$, $\lambda > 0$ the curve

$$\tilde{\gamma}(t) := \gamma(\lambda t + t_0), \qquad \frac{a - t_0}{\lambda} < t < \frac{b - t_0}{\lambda}$$

is still in G;

The pair $(M, \mathcal{G})$ is called a path space. For a vector $v \in T_xM$, let $\gamma : (-\varepsilon, \varepsilon) \to M$ be the curve in $\mathcal{G}$ with $\dot{\gamma}(0) = v$. Define

$$G^i(v) := -\frac{1}{2}\frac{d^2\gamma^i}{dt^2}(0), \tag{4.3}$$

where $(\gamma^i(t))$ are the local coordinates of the curve $\gamma(t)$ and $G^i(v)$ are positive homogeneous of degree 2 in v. It is then easy to prove that any curve $c : (a, b) \to M$

in $\mathcal{G}$ satisfies the following system,

$$\frac{d^2\gamma^i}{dt^2} + 2G^i\left(\frac{dc}{dt}\right) = 0. \tag{4.4}$$

Coordinate Transformation

Consider local coordinate changes on M, say

$$x^i = x^i(\tilde{x}^1, \dots, \tilde{x}^n),$$

and their inverses $\tilde{x}^p = \tilde{x}^p(x^1, \dots, x^n)$. They induce transformations on the manifold TM, as follows: observe that the tangent bundle of the manifold TM has a local coordinate basis that consists of the bases $\frac{\partial}{\partial x^j}$ and $\frac{\partial}{\partial v^j}$. However, under transformation on TM induced by a coordinate change on M, the vectors $\frac{\partial}{\partial x^j}$ transform in a somewhat complicated manner,

$$\frac{\partial}{\partial \tilde{x}^p} = \frac{\partial x^i}{\partial \tilde{x}^p}\frac{\partial}{\partial x^i} + \frac{\partial^2 x^i}{\partial \tilde{x}^p \partial \tilde{x}^q}\tilde{v}^q\frac{\partial}{\partial v^i} \tag{4.5}$$

The second term on the right hand side of (4.5) is not tensorial. On the other hand $\frac{\partial}{\partial v^j}$ transform ordinarily

$$\frac{\partial}{\partial \tilde{v}^p} = \frac{\partial x^i}{\partial \tilde{x}^p}\frac{\partial}{\partial v^i}. \tag{4.6}$$

Similarly, the cotangent bundle of TM has a local coordinate basis $\{dx^i, dv^i\}$. Here, under a coordinate transformation, the dx^i behave simply while the dv^i do not.

$$d\tilde{x}^p = \frac{\partial \tilde{x}^p}{\partial x^j} dx^i \tag{4.7}$$

$$d\tilde{v}^p = \frac{\partial \tilde{x}^p}{\partial x^i} dv^i + \frac{\partial^2 \tilde{x}^p}{\partial x^i \partial x^j} v^j dx^i \tag{4.8}$$

The remedy lies in replacing $\frac{\partial}{\partial x^i}$ by

$$\frac{\delta}{\delta x^j} \ := \ \frac{\partial}{\partial x^j} - N^i_j \frac{\partial}{\partial v^i} \tag{4.9}$$

and dv^i by

$$\delta v^i \ := \ dv^i + N^i_j dx^j, \tag{4.10}$$

where

$$N^i_j(x, v) \ := \ \Gamma^i_{jk} v^k = \frac{\partial G^i}{\partial v^i}(x, v).$$

Its clear that the objects $\frac{\delta}{\delta x^j}$ and dx^j are the natural duals of each other, similarly, the objects $\frac{\partial}{\partial v^i}$ and δv^i are dual to each other. All these objects lie on $TM \setminus \{0\}$. While dx^i is holonomic, the rest are non-holonomic. We now show how these objects split the tangent and cotangent spaces of $TM \setminus \{0\}$ and introduce a nonlinear connection on it.

Decomposition of $T(TM \setminus \{0\})$ and $T^*(TM \setminus \{0\})$

Let M be an n-dimensional manifold and $TM \setminus \{0\}$ denote the slit tangent bundle. Take a standard local coordinate system (x^i, v^i) in TM. Let $\{\frac{\delta}{\delta x^i}, \frac{\partial}{\partial y^i}\}$ denote the local frame dual to $\{dx^i, \delta y^i\}$, that is

$$\frac{\delta}{\delta x^i} := \frac{\partial}{\partial x^i} - N_i^j(x,v)\frac{\partial}{\partial v^j}. \tag{4.11}$$

Let

$$\mathcal{V}TM := \mathsf{span}\left\{\frac{\partial}{\partial v^1}, \ldots, \frac{\partial}{\partial v^n}\right\} \tag{4.12}$$

$\mathcal{V}TM$ is a well-defined sub-bundle of $T(TM \setminus \{0\})$. We call $\mathcal{V}TM$ the vertical tangent bundle over $TM \setminus \{0\}$. There is a canonical vector field on $TM \setminus \{0\}$ defined by

$$Y := v^i\frac{\partial}{\partial v^i}. \tag{4.13}$$

Y is called the vertical radial field which is a section of $\mathcal{V}TM$. Given a spray on M,

$$\mathbb{G} = v^i\frac{\partial}{\partial x^i} - 2G^i\frac{\partial}{\partial v^i}. \tag{4.14}$$

Let

$$N_j^i := \frac{\partial G^i}{\partial v^j}. \tag{4.15}$$

We call N_j^i the connection coefficients of G. The homogeneity condition (4.2) implies

$$N_j^i(x,\lambda v) = \lambda N_j^i(x,v), \qquad \forall\lambda > 0, \tag{4.16}$$

and

$$G^i(x,v) = N_j^i(x,v)v^j. \tag{4.17}$$

Let

$$\mathcal{H}TM := \mathsf{span}\left\{\frac{\delta}{\delta x^1}, \ldots, \frac{\delta}{\delta x^n}\right\} \tag{4.18}$$

where

$$\frac{\delta}{\delta x^j} \ := \ \frac{\partial}{\partial x^j} - N^i_j \frac{\partial}{\partial v^i} \tag{4.19}$$

$\mathcal{H}TM$ is a well-defined sub-bundle of $T(TM \setminus \{0\})$. We call $\mathcal{H}TM$ the horizontal tangent bundle over $TM \setminus \{0\}$. It follows from (4.19) that

$$\mathbb{G} = v^i \frac{\delta}{\delta x^i}. \tag{4.20}$$

Therefore G is a horizontal vector field which is a section of $\mathcal{H}TM$. We therefore obtain a direct decomposition of the tangent bundle of $T(TM \setminus 0)$,

$$T(TM \setminus \{0\}) = \mathcal{H}TM \oplus \mathcal{V}TM. \tag{4.21}$$

Let

$$\mathcal{H}^*TM := \mathsf{span}\ \left\{dx^1, \dots, dx^n\right\} \tag{4.22}$$

and

$$\mathcal{V}^*TM := \mathsf{span}\ \left\{\delta v^1, \dots, \delta v^{in}\right\} \tag{4.23}$$

where

$$\delta v^i \ := \ dv^i + N^i_j dx^j \tag{4.24}$$

HTM and VTM are well defined sub-bundles of $T^*(TM \setminus \{0\})$. We obtain a decomposition for $T^*(TM \setminus 0)$, which is dual to (4.21),

$$T^*(TM \setminus \{0\}) \ = \mathcal{H}^*TM \oplus \mathcal{V}^*TM. \tag{4.25}$$

Given a spray $\mathbb{G}$ on a manifold M. We can construct lots of sprays from $\mathbb{G}$ using the vertical radial field Y. Let $\mathbb{Q} = Q^i(v)\frac{\partial}{\partial v^i}$ be a vertical vector field on $TM \setminus \{0\}$ which has the following properties:

1. $\mathbb{Q}$ is C^∞ on $TM \setminus \{0\}$,

2. the coefficients $Q^i(v)$ are positively homogeneous of degree two, i.e., $Q^i(\lambda v) = \lambda^2 Q^i(v)$, $\lambda > 0$.

Defining $\tilde{\mathbb{G}}$ as

$$\tilde{\mathbb{G}} = \mathbb{G} - 2\mathbb{Q}, \tag{4.26}$$

then $\tilde{\mathbb{G}}$ is also a spray. Furthermore, $\mathbb{G}$ and $\tilde{\mathbb{G}}$ have the same geodesics as point sets. That is every geodesic of $\mathbb{G}$, after reparametrization, becomes a geodesic of $\tilde{\mathbb{G}}$, and vice versa.

4.2.1 Finsler Sprays

When the spray

$$\mathbb{G} := v^i \frac{\partial}{\partial x^i} - 2G^i(x,v)\frac{\partial}{\partial v^i} \tag{4.27}$$

is induced by a Finsler metric F we call them Finsler sprays. As shown in a previous chapter that the geodesic coefficients G^i are given by

$$G^i(v) := \frac{1}{4} g^{il} \left[F^2_{v^l x^k} v^k - F^2_{x^l} \right] \tag{4.28}$$

and from the homogeneity of F, we have

$$G^i(x, \lambda v) = \lambda^2 \, G^i(x,v), \qquad \lambda > 0. \tag{4.29}$$

The geodesics of $\mathbb{G}$ are called the geodesics of F. The spray coefficients G^i of $\mathbb{G}$ are also called the spray coefficients of F.

Let $\mathbb{G}$ be a spray induced by the metric F, then the connection coefficients N^i_j and the Christoffel symbols Γ^i_{jk} are given by

$$N^i_j(x,v) := \frac{\partial G^i}{\partial v^j}(x,v), \tag{4.30}$$

and

$$\Gamma^i_{jk} := \frac{\partial^2 G^i}{\partial v^j \partial v^k}(x,v). \tag{4.31}$$

The homogeneity of G^i, (4.2), then implies that

$$N^i_j = \Gamma jk^i(x,v)v^k, \quad \text{and} \quad 2G^i(x,v) = N^i_j(x,v)v^j = \Gamma^i_{jk}(x,v)v^jv^k. \tag{4.32}$$

Lemma 4.2.3. *For any Finsler metric F on a manifold*

$$\frac{\delta}{\delta x^k}F = 0. \tag{4.33}$$

Hence

$$\mathbb{G}(F) = v^k\frac{\delta}{\delta x^k}F = 0. \tag{4.34}$$

For a Finsler metric F on a manifold M, there is a canonical 1-form ω on $TM \setminus \{0\}$, defined by

$$\omega := g_{ij}(x,v)v^j dx^i. \tag{4.35}$$

The form ω is called the Hilbert form. Furthermore, let

$$\theta := g_{ij}(x,v)dx^i \wedge \delta v^j. \tag{4.36}$$

We then obtain the following lemma.

Lemma 4.2.4. *The Hilbert form* ω *and the Finsler spray* $\mathbb{G}$ *are related by*

$$\omega(\mathbb{G}) = F, \qquad \theta(\mathbb{G}, \cdot) = \frac{1}{2}dF. \tag{4.37}$$

4.3 Non-Riemannian quantities

In Riemannian geometry the main geometric objects are the Riemannian curvature and sectional curvature. However in Finsler geometry and the geometry of sprays there are additional geometric objects which vanish in the Riemannian case. We now describe these quantities which will be helpful in the derivation of connections.

4.3.1 Cartan Torsion

Let (M, F) be a Finsler manifold. For a vector $y \in T_xM \setminus \{0\}$, let

$$C_{ijk}(y) \;:=\; \frac{1}{4}[F^2]_{y^iy^jy^k}(y) = \frac{\partial g_{ij}}{\partial y^k}(y) \tag{4.38}$$

Define the map $\mathbf{C}_y : T_xM \otimes T_xM \otimes T_xM \to R$ by

$$C_y(u, v, w) \;=\; C_{ijk}(y)u^iv^jw^k, \tag{4.39}$$

For each $v \in T_xM \setminus \{0\}$, C_v is a symmetric multi-linear form on $T_xM \setminus \{0\}$. The homogeneity of F implies

$$\mathbf{C}_v(v, a, b) \;=\; 0, \qquad a, b \in T_xM \setminus \{0\}. \tag{4.40}$$

The family $\mathbf{C} = \{\mathbf{C}_v\}_{v \in T_xM\setminus\{0\}}$ is called the *Cartan torsion*. It is a trivial result that the Cartan torsion vanishes if and only if F^2 is quadratic.

The mean Cartan torsion is defined by the following map, $\mathbf{I}_y : T_xM \to R$ by

$$\mathbf{I}_y(u) \;=\; I_i(y)u^i \tag{4.41}$$

where

$$I_i(y) \;:=\; g^{ik}(y)C_{ijk}(y). \tag{4.42}$$

It again follows from homogeneity that

$$\mathbf{I}_y(y) \;=\; 0. \tag{4.43}$$

4.3.2 Berwald Curvature

Upon change in coordinate system from (x^i) to $(\tilde{x}^i)$, the Γ^i_{jk} change according to

$$\tilde{\Gamma}^p_{qr} = \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial x^k}{\partial \tilde{x}^r}\Gamma^i_{jk} + \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^i}{\partial \tilde{x}^q \partial \tilde{x}^r} \tag{4.44}$$

It is easy to see that the Γ^i_{jk}'s are not tensors, however, upon differentiating (4.44) with respect to v^l, we have

$$\frac{\partial \tilde{\Gamma}^p_{qr}}{\partial \tilde{v}^s}\frac{\partial \tilde{x}^s}{\partial x^l} = \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial x^k}{\partial \tilde{x}^r}\frac{\partial \Gamma^i_{jk}}{\partial v^l} \tag{4.45}$$

We now see that $\frac{\partial \Gamma^i_{jk}}{\partial v^l}$ are the coefficients of a tensor on $TM \setminus \{0\}$.

Setting

$$B^i_{jkl}(x,v) := \frac{\Gamma^i_{jk}}{\partial v^l}(x,v) = \frac{\partial^3 G^i}{\partial v^j \partial v^k \partial v^l}(x,v) \tag{4.46}$$

we now define a new geometric quantity called Berwald curvature. It was Berwald who first noticed that the third order derivatives of G^i give rise to an invariant of

Finsler sprays. For a tangent vector $v \in T_xM \setminus \{0\}$, define

$$\mathbf{B}_y : T_xM \otimes T_xM \otimes T_xM \to T_xM \tag{4.47}$$

by

$$\mathbf{B}_y(u,v,w) := B^i_{jkl}(x,v)u^j v^k w^l \frac{\partial}{\partial x^i}|_x, \tag{4.48}$$

$\mathbf{B}_y(u,v,w)$ is symmetric in u, v, and w. Furthermore, the homogeneity of G^i implies

$$\mathbf{B}_y(y,v,w) = 0. \tag{4.49}$$

Let

$$E_{jk} := \frac{1}{2}B^m_{jkm}(x,v) = \frac{1}{2}\frac{\Gamma^m_{jk}}{\partial v^m}(x,v) = \frac{1}{2}\frac{\partial^3 G^m}{\partial v^j \partial v^k \partial v^m}(x,v) \tag{4.50}$$

This set of local functions give rise to a tensor on $TM \setminus \{0\}$. For a tangent vector $v \in T_xM \setminus \{0\}$, define

$$\mathbf{E}_y : T_xM \otimes T_xM \to R \tag{4.51}$$

by

$$\mathbf{E}_y(u,v) := E_{jk}(y)u^j v^k. \tag{4.52}$$

$\mathbf{E}$ is called the mean Berwald curvature. $\mathbf{E}_y(u,v)$ is symmetric in u and v and can be viewed as the trace of $\mathbf{B}_y$. It further follows from homogeneity of spray coefficient that

$$\mathbf{E}_y(y,v) = 0. \tag{4.53}$$

4.3.3 Landsberg Curvature

Let (M, F) be a Finsler space. For a tangent vector $y \in TM \setminus \{0\}$, define

$$\mathbf{L}_y(u, v, w) := -\frac{1}{2} g_y(B_y(u, v, w), y). \tag{4.54}$$

In local coordinates

$$\mathbf{L}_y(u, v, w) := L_{ijk}(x, y) u^i v^j w^k \tag{4.55}$$

where $u = u^i \frac{\partial}{\partial x^i}\big|_x, u = u^i \frac{\partial}{\partial x^i}\big|_x, u = u^i \frac{\partial}{\partial x^i}\big|_x \in T_x M$ and

$$L_{ijk}(y) := -\frac{1}{2} y^m g_{ml}(y) B^l_{ijk}(y) = -\frac{1}{2} y^m g_{ml}(y) \frac{\partial^3 G^l}{\partial y^i \partial y^j \partial y^k}. \tag{4.56}$$

$\mathbf{L} := \{\mathbf{L}_v\}_{v \in T_x M \setminus \{0\}}$ is called the Landsberg curvature. $\mathbf{L}_v$ is a symmetric multilinear form. It

$$\mathbf{L}_y(y, v, w) \ = \ 0. \tag{4.57}$$

Let $c(t)$ be an arbitrary geodesic in (M, F). Take arbitrary parallel vector fields $U(t), V(t), W(t)$ along c. We then have

$$\mathbf{L}_{\dot{c}(t)}(U, V, W) = \frac{d}{dt} \mathbf{C}_{\dot{c}(t)}(U, V, W). \tag{4.58}$$

Thus the Landsberg curvature measures the rate of change of the Cartan torsion along geodesics.

4.4 Connections

We first define the general notion of connections, affine connections, and the torsion-freeness condition.

Definition 4.4.1. *A connection ∇ on M is a family of maps*

$$\nabla := \{\nabla^y : T_xM \times C^\infty(TM) \to T_xM, \qquad y \in T_xM \setminus \{0\},\ x \in M\}$$

which has the following properties

1. $\nabla_u^{\lambda y} V = \nabla_u^y V$;
2. $\nabla_u^y (fV) = u(f)V + f\nabla_u^y V$;
3. $\nabla_u^y (U + V) = \nabla_u^y U + \nabla_u^y V$;
4. $\nabla_{fu}^y V = f\nabla_u^y V$;
5. $\nabla_{u+v}^y V = \nabla_u^y V + \nabla_v^y V$;
6. $\nabla_U^Y V - \nabla_V^Y U = [U, V]$;

where $\lambda > 0$, $u, v \in T_xM$ and $Y, U, V \in C^\infty(TM)$.

Let ∇ be a connection on a manifold M. In a standard local coordinate system (x^i, v^i) in TM, define a set of local functions $\Gamma^i_{jk}(v)$ on TM by

$$\Gamma^i_{jk}(v) \frac{\partial}{\partial x^i}\bigg|_x := \nabla^y_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^i}, \qquad y \in T_xM. \tag{4.59}$$

Property (6) in Definition 4.4.1 implies

$$\Gamma^i_{jk} = \Gamma^i_{kj}. \tag{4.60}$$

In this sense, ∇ is torsion-free. For any $u = u^i \frac{\partial}{\partial x^i}\Big|_x$ and $V = V^i \frac{\partial}{\partial x^i}$,

$$\nabla_u^y V = \{u(V^i)(x) + V^j(x)\Gamma^i_{jk}(y)u^k\} \frac{\partial}{\partial x^i}\bigg|_x. \tag{4.61}$$

When ∇^y is independent of $y \in TM \setminus \{0\}$, ∇ is called an affine connection.

4.4.1 Berwald Connection

There is a canonical connection for every spray. Let

$$\mathbb{G} = v^i \frac{\partial}{\partial x^i} - 2G^i(x,v)\frac{\partial}{\partial v^i} \tag{4.62}$$

be a spray on a manifold M. The Christoffel symbols of G are defined by

$$\Gamma^i_{jk}(v) \ := \ \frac{\partial^2 G^i}{\partial v^j \partial v^k}(v). \tag{4.63}$$

The Berwald connection is defined by the map $\nabla^y : T_xM \times C^\infty(TM) \to T_xM$ given by

$$\nabla^y_u V \ = \ \left\{u(V^i)(x) + V^j(x)\Gamma^i_{jk}(y)u^k\right\} \frac{\partial}{\partial x^i}\bigg|_x \tag{4.64}$$

Clearly this connection is torsion-free.

On a Finsler space $(M.F)$, the Berwald connection ∇ is the unique connection on TM satisfying

$$\nabla^Y_U V - \nabla^Y_V U = [U,V] \tag{4.65}$$

$$\begin{aligned} W[g_Y(U,V)] - g_Y(\nabla^Y_W U, V) - g_Y(U, \nabla^Y_W V) &= 2C_Y(U,V,\nabla^Y_W Y) \\ &\quad - 2L_Y(U,V,W) \end{aligned} \tag{4.66}$$

where $U,Y,V,W \in C^\infty(TM)$. Equation (4.65) is equivalent to torsion freeness.

4.4.2 Chern connection

The Chern connection differs from the Berwald connection only by a term called Landsberg curvature. Let (M,F) be a Finsler space. For a vector $y \in T_xM \setminus \{0\}$, the Landsberg curvature $\mathbf{L}_y : T_xM \times T_xM \times T_xM \to R$ determines a bilinear symmetric form $\mathbf{L}_y : T_xM \otimes T_xM \to T_xM$ by

$$g_y(L_y(u,v),w) = L_y(u,v,w). \tag{4.67}$$

and by (4.57), it follows that

$$\mathbf{L}_y(y,v) = 0 \tag{4.68}$$

The Chern connection ${}^C\nabla$ is defined from the Berwald connection ∇ as follows: define a map ${}^C\nabla^y : T_xM \times C^\infty(TM) \to T_xM$ by

$${}^C\nabla^y_u V := \nabla^y_u V - \mathbf{L}_y(u,v), \tag{4.69}$$

where $u, v \in T_xM$ and $V \in C^\infty(TM)$ with $V_x = v$. ${}^C\nabla = \{{}^C\nabla^y\}_{y \in TM\setminus\{0\}}$ is called the Chern connection.

In a standard local coordinate system (x^i, v^i) in TM, let

$${}^C\Gamma^i_{jk}(v) := \Gamma^i_{jk}(v) - L^i_{jk}(v), \tag{4.70}$$

where $L^i_{jk} := g^{il}L_{jkl}$, then the Chern connection can be expressed as

$${}^C\nabla^y_u V = \left\{u(V^i)(x) + V^j\, {}^C\Gamma^i_{jk}(y)u^k\right\} \frac{\partial}{\partial x^i}\bigg|_x. \tag{4.71}$$

The Chern connection is characterized by the following equations

$${}^C\nabla^Y_U V - {}^C\nabla^Y_V U = [U,V], \tag{4.72}$$

$$w[g_Y(U,V)] - g_Y({}^C\nabla^Y_w U, V) - g_Y(U, {}^C\nabla^Y_w V) = C_Y(U,V,{}^C\nabla^Y_w Y), \tag{4.73}$$

where $w \in T_xM$ and $U, V, Y \in C^\infty(TM)$. (4.72) is the torsion-freeness condition and (4.73) is almost metric-compatibility condition. In the differential forms notation the Chern connection can be written as shown in the following theorem:

Theorem 4.4.2. *Let (M, F) be a Finsler manifold. The pulled-back bundle π^*TM admits a unique linear connection ${}^C\nabla_v$, called the Chern connection. Its connection forms ω_j^i are characterized by the structural equations:*

$$d(dx^i) = dx^j \wedge \omega_j^i, \tag{4.74}$$

$$dg_{ij} = g_{kj}\omega_i^k + g_{kj}\omega_i^k + 2A_{ijk}\frac{\delta v^k}{F}. \tag{4.75}$$

Torsion freeness is equivalent to the absence of dv^k terms in ω_j^i; namely

$$\omega_j^i = {}^C\Gamma^i_{jk} dx^k, \tag{4.76}$$

together with the symmetry

$${}^C\Gamma^i_{kj} = {}^C\Gamma^i_{jk}. \tag{4.77}$$

Almost metric-compatibility then implies that

$${}^C\Gamma^i_{jk} = \Gamma^i_{jk} - g^{il}\left(A_{ljs}\frac{N_k^s}{F} - A_{jks}\frac{N_l^s}{F} + A_{kls}\frac{N_j^s}{F}\right) \tag{4.78}$$

which could be written as

$${}^C\Gamma^i_{jk} = g^{is}\left(\frac{\delta g_{sj}}{\delta x^k} - \frac{\delta g_{jk}}{\delta x^{ks}} + \frac{\delta g_{ks}}{\delta x^j}\right). \tag{4.79}$$

Chern Curvature

The Chern connection gives rise to the following quantity

$$P^i_{jkl} := -\frac{\partial^C \Gamma^i_{jk}}{\partial v^l} \;=\; -\frac{\partial^3 G^i}{\partial v^j \partial v^k \partial v^l} + \frac{\partial L^i_{jk}}{\partial v^l}. \tag{4.80}$$

Note that

$$P^i_{jkl} y^l = 0, \qquad P^i_{jkl} y^j = -L^i_{kl}. \tag{4.81}$$

For a vector $y \in T_xM \setminus \{0\}$, define $\mathbf{P}_y : T_xM \otimes T_xM \otimes T_xM \to T_xM$ by

$$\mathbf{P}_y(u, v, w) := P^i_{jkl}(x, y) u^j v^k w^l \frac{\partial}{\partial x^i}\Big|_x \tag{4.82}$$

and $\mathbf{P} := \{\mathbf{P}_v\}_{v \in T_xM \setminus \{0\}}$ is called the Chern curvature.

4.4.3 Covariant Derivatives Along Geodesics

Let G be a spray on an n-manifold and Γ^i_{jk} the Christoffel symbols of G in a standard local coordinates on TM. Let $c : [a, b] \to M$ be a C^∞ curve. A vector field $V = V(t)$ along c is a family of vectors

$$V(t) = V^i(t) \frac{\partial}{\partial x^i}\Big|_{c(t)},$$

where $V^i(t)$ are C^∞. Hence the tangent vector field of c,

$$\dot{c}(t) = \frac{dc^i}{dt}(t) \frac{\partial}{\partial x^i}\Big|_{c(t)},$$

is a vector field along c.

By fixing a vector field $V = V(t)$ along c and given a vector field $U(t)$ along c, we can define the following connection along c:

$$\nabla^V_{\dot{c}(t)} U(t) \;:=\; \left\{ \frac{dU^i}{dt}(t) + U^j \Gamma^i_{jk}(V(t)) \frac{dc^k}{dt}(t) \right\} \frac{\partial}{\partial x^i}\bigg|_{c(t)}. \qquad (4.83)$$

At a point $c(t_0)$ where $\dot{c}(t_0) = 0$, reduces to

$$\nabla^V_{\dot{c}} U(t_0) \;:=\; \frac{dU^i}{dt}(t_0) \frac{\partial}{\partial x^i}\bigg|_{c(t_0)} \qquad (4.84)$$

which does not depend on the vector field V.

When the vector field $V(t)$ is the coincides with the $\dot{c}(t)$ along the curve c, then we define the differentiation of the vector U along c by

$$D_{\dot{c}} U(t) := \nabla^{\dot{c}}_{\dot{c}} U(t), \qquad (4.85)$$

$$= \left\{ \frac{dU^i}{dt}(t) + U^j \Gamma^i_{jk}(\dot{c}(t)) \frac{dc^k}{dt}(t) \right\} \frac{\partial}{\partial x^i}\bigg|_{c(t)}. \qquad (4.86)$$

$D_{\dot{c}} U(t)$ is called the covariant derivative of $U(t)$ along the curve c.

Given the following optimal control problem

$$\min_{u \in \mathcal{U}} \int_{t_0}^{t_f} L(x, u) dt \qquad (4.87)$$

subject to the control system

$$\dot{x} = f(x, u) \qquad (4.88)$$

the geodesic spray in the tangent space $(\mathrm{x}, \mathrm{x}') = (\mathrm{x}, (\mathrm{t}', \mathrm{y}', \mathrm{c}, \mathrm{p}))$ is given by

$$\mathbb{G} = \mathrm{x}' \frac{\partial}{\partial \mathrm{x}} + {}^{\mathrm{t}}G^i(\mathrm{x}, \mathrm{x}') \frac{\partial}{\partial \mathrm{t}'} + {}^{\mathrm{y}}G^i(\mathrm{x}, \mathrm{x}') \frac{\partial}{\partial \mathrm{y}'^i} + {}^{\mathrm{c}}G^i(\mathrm{x}, \mathrm{x}') \frac{\partial}{\partial \mathrm{c}^i} + {}^{\mathrm{p}}G^i(\mathrm{x}, \mathrm{x}') \frac{\partial}{\partial \mathrm{p}^i}. \qquad (4.89)$$

The covariant derivative for the spray (4.89) is then defined using the Chern connection to be

$$D_u^{f,L} V := \left\{ u(V^i)(x) + V^j \, \frac{\partial G^i}{\partial \mathrm{x}'^j \partial \mathrm{x}'^k}(u) u^k - V^j L^i_{jk}(u) u^k \right\} \frac{\partial}{\partial x^i} \tag{4.90}$$

From these connections one can define the covariant derivatives and parallel transport equations as follows.

4.4.4 Parallel Transport

Along the curve c the covariant derivative of its tangent vector $\dot{c}$ is given by

$$D_{\dot{c}} \dot{c} = \left\{ \ddot{c}^i + \dot{c}(t) N^i_j(\dot{c}(t)) \right\} \left. \frac{\partial}{\partial x^i} \right|_{c(t)}, \tag{4.91}$$

$$= \left\{ \ddot{c}^i + 2G^i(\dot{c}) \right\} \left. \frac{\partial}{\partial x^i} \right|_{c(t)} \tag{4.92}$$

It is quite obvious that c is a geodesic if and only if

$$D_{\dot{c}} \dot{c} \ = \ 0. \tag{4.93}$$

Definition 4.4.3. *A vector field $V = V(t)$ along c is said to be parallel if*

$$D_{\dot{c}} V \ = \ 0. \tag{4.94}$$

Let $p = c(a)$ and $q = c(b)$, then the parallel transport $P_c : T_x M \to T_x M$ of a vector $V(t) \in T_{c(t)} M$ along the curve c is defined by

$$P_c(v) := V(b), \qquad v \in T_p M \tag{4.95}$$

where $V(t)$ is a parallel vector field along c with $V(a) = v$. This map is a linear map.

In the context of the optimal control problems, parallel transport of a vector V along the curve c is given by the equation

$$D^{f,L}_{\dot{c}} V = 0 \tag{4.96}$$

where

$$D^{f,L}_{\dot{c}} V = \left\{ \dot{c}(V^i)(x) + V^j \, \frac{\partial G^i}{\partial \mathrm{x}'^j \partial \mathrm{x}'^k}(\dot{c}) \frac{dc^k}{dt} - V^j L^i_{jk}(\dot{c}) \frac{dc^k}{dt} \right\} \frac{\partial}{\partial x^i} \tag{4.97}$$

In the next chapter, we use these connections to derive the curvature operators which determine the spread of geodesics for small time.

Chapter 5

Curvatures and the Spread of Optimal Controls

In order to derive control laws in presence of disturbances one needs to derive controls as a function of state, $u(x)$, which are usually called feedback controls. The synthesis problem for optimal controls remains a principal problem. However, if one assumes that the disturbances in the system dynamics are small in magnitude and bounded then one should be able to derive those optimal controls as a function of those disturbances of state. In this chapter we deal with this problem of synthesizing controls which takes into account the geometry of the given optimal control problem.

In the previous chapters we have found that, given an optimal control problem one could devise a Finsler metric. We further defined the notion of an affine connection on this Finsler space (M, g) which gives a way of differentiating vector fields on this manifold along a given direction. We now show how this connection gives us certain geometric properties of the manifold in terms of torsion and curvature which are third and fourth order tensors respectively.

On Riemannian manifolds there exists an unique affine (Levi-Civita)connection which is torsion free and metric compatible. The curvature tensor which is an intrinsic object of the manifold helps in studying geodesic flows. In Euclidean spaces the curvature term always vanishes since the derivatives $\partial_i \partial_j$ always commute, i.e.,

$$\frac{\partial^2}{\partial x^i \partial x^j} = \frac{\partial^2}{\partial x^j \partial x^i}.$$

However this is not the case on Riemannian manifolds, as the commutator of covariant derivatives defined by the Levi-Civita connection, $[D_i, D_j]$ does not vanish and is given by the following index free notation

$$R(U, V)W = (D_U D_V - D_V D_U - D_{[U,V]})W, \tag{5.1}$$

where U, V, and W are vector fields on M.

In Finsler geometry the curvature tensor not only depends on the position but also the direction. Here we derive the curvature tensors from the Chern connection using Cartan's structural equation. We do not have just one curvature but have two different objects known in literature as Riemann and Chern curvatures. The Riemann curvature gives us a measure of the the dispersion of the geodesics with different initial conditions. The Chern curvature can be derived from Cartan's torsion, which is defined in Minkowski space. Since every tangent space on a Finsler space is a Minkowski space, Cartan torsion is defined for a Finsler space. We outline the derivation of these two curvature tensors in this chapter and study how they affect the spread of the optimal controls for neighboring points on the manifold.

5.1 Curvature

5.1.1 Curvatures from the Chern Connection

We now use exterior differential algebra to derive curvature quantities from the connection. The Chern connection is uniquely determined by

$$d\omega^i = \omega^j \wedge \omega_j^i \tag{5.2}$$

$$dg_{ij} = g_{kj}\omega_i^k + g_{ik}\omega_j^k + 2A_{ijs}\frac{\delta v^s}{F}. \tag{5.3}$$

The curvature 2-forms of the Chern connection Ω_j^i are defined as

$$\Omega_j^i := d\omega_j^i - \omega_j^k \wedge \omega_k^i. \tag{5.4}$$

Since $\{\Omega_j^i\}$ are a set of local 2-forms on $TM \setminus \{0\}$, they can be expanded in the following form

$$\Omega_j^i = \frac{1}{2}R_{jkl}^i dx^k \wedge dx^l + P_{jkl}^i dx^k \wedge \frac{\delta v^l}{F} + \frac{1}{2}Q_{jkl}^i \frac{\delta v^k}{F} \wedge \frac{\delta v^l}{F}, \tag{5.5}$$

where one can assume that the following relations hold

$$R^i_{jkl} + R^i_{jlk} = 0 \tag{5.6}$$

$$Q^i_{jkl} + Q^i_{jlk} = 0. \tag{5.7}$$

Exterior differentiation of the torsion freeness condition $dx^j \wedge \omega^i_j = 0$ gives

$$dx^j \wedge d\omega^i_j = 0. \tag{5.8}$$

Since the term $dx^j \wedge \omega^k_j \wedge \omega^i_k$ vanishes by torsion freeness, it can be subtracted from the left hand side of equation (5.8) to obtain the following equation

$$\omega^j \wedge \Omega^i_j = 0. \tag{5.9}$$

Substituting equation (5.5) for Ω^i_j into the above equation we get

$$\frac{1}{2}R^i_{jkl}\, dx^j \wedge dx^k \wedge dx^l + P^i_{jkl}\, dx^j \wedge dx^k \wedge \frac{\delta v^l}{F} + \frac{1}{2}Q^i_{jkl}\, dx^j \wedge \frac{\delta v^k}{F} \wedge \frac{\delta v^l}{F} = 0, \tag{5.10}$$

and by the antisymmetric properties of the 3-forms in the above equation we have

$$R^i_{jkl} + R^i_{klj} + R^i_{ljk} = 0, \tag{5.11}$$

$$P_{jkl} = P^i_{kjl}, \tag{5.12}$$

$$Q^i_{jkl} = 0. \tag{5.13}$$

The curvature 2-form (5.5) then simplifies to

$$\Omega^i_j = \frac{1}{2}R^i_{jkl}dx^k \wedge dx^l + P^i_{jkl}dx^k \wedge \frac{\delta v^l}{F}. \tag{5.14}$$

5.1.2 Formula for R and P in natural coordinates

By Chern's connection(torsion freeness) we have,

$$\omega_j^i = \Gamma^i_{jk} dx^k = \left(\frac{\partial^2 G^i}{\partial v^j \partial v^k} - L^i_{jk} \right) dx^k.$$

from which it follows that

$$d\omega_j^i = d\Gamma^i_{jk} \wedge dx^k \tag{5.15}$$

Since the differential $d\Gamma^i_{jk}$ is a 1-form on $TM \setminus \{0\}$, it can be expanded in terms of dx^k and $\frac{\delta v^k}{F}$ as follows

$$d\omega_j^i = \frac{\delta \Gamma^i_{jl}}{\delta x^k}\, dx^k \wedge dx^l - F \frac{\partial \Gamma^i_{jl}}{\partial v^k}\, dx^k \wedge \frac{\delta v^l}{F}. \tag{5.16}$$

Furthermore we have

$$\begin{aligned} -\omega_j^h \wedge \omega_h^i &= \omega_h^i \wedge \omega_j^h \\ &= \Gamma^i_{hk}\, \Gamma^h_{jl}\, dx^k \wedge dx^l. \end{aligned} \tag{5.17}$$

Substituting equations (5.16) and (5.17) into Ω^i_j, we have

$$\Omega = \left(\frac{\delta \Gamma^i_{jl}}{\delta x^k} + \Gamma^i_{hk}\, \Gamma^h_{jl} \right) dx^k \wedge dx^l - F \frac{\partial \Gamma^i_{jk}}{\partial v^l}\, dx^k \wedge \frac{\delta v^l}{F}. \tag{5.18}$$

Equating (5.14) and (5.18), we obtain

$$R^i_{jkl} = \frac{\delta\Gamma^i_{jl}}{\partial x^k} - \frac{\delta\Gamma^i_{jk}}{\partial x^l} + \Gamma^m_{jl}\,\Gamma^i_{mk} - \Gamma^m_{jk}\,\Gamma^i_{ml} \tag{5.19}$$

$$= \frac{\partial\Gamma^i_{jl}}{\partial x^k} - \frac{\partial\Gamma^i_{jk}}{\partial x^l} + \frac{\partial\Gamma^i_{jk}}{\partial v^m}N^m_l - \frac{\partial\Gamma^i_{jl}}{\partial v^m}N^m_k + \Gamma^m_{jl}\,\Gamma^i_{mk} - \Gamma^m_{jk}\,\Gamma^i_{ml} \tag{5.20}$$

$$P^i_{jkl} = -\frac{\partial\Gamma^i_{jk}}{\partial v^l} \;=\; -\frac{\partial^3 G^i}{\partial v^j \partial v^k \partial v^l} + \frac{\partial L^i_{jk}}{\partial v^l} \tag{5.21}$$

By homogeneity of F, we obtain

$$R^i_k = 2\frac{\partial G^i}{\partial x^k} - v^j\frac{\partial^2 G^i}{\partial x^j \partial v^k} + 2G^j\frac{\partial^2 G^i}{\partial v^j \partial v^k} - \frac{\partial G^i}{\partial v^j}\frac{\partial G^j}{\partial v^k} \tag{5.22}$$

$$P^i_{kl} = -L^i_{kl} \tag{5.23}$$

Exterior differentiation of the almost metric compatibility of the Chern connection

$$dg_{ij} - g_{kj}\omega^k_i - g_{ik}\omega^k_j = 2A_{ijs}\frac{\delta v^s}{F} \tag{5.24}$$

we have

$$\Omega_{ij} + \Omega_{ji} = \frac{1}{2}(R_{ijkl} + R_{jikl})dx^k \wedge dx^l + (P_{ijkl} + P_{jikl})dx^k \wedge \frac{\delta v^l}{F} \tag{5.25}$$

$$= -(A_{iju}R^u_{kl})dx^k \wedge dx^l - 2(A_{iju}P^u_{kl} + A_{ijl|k})dx^k \wedge \frac{\delta v^l}{F}$$
$$+ 2(A_{ijk;l} - A_{ijk}\ell_l)\frac{\delta v^k}{F} \wedge \frac{\delta v^l}{F} \tag{5.26}$$

The following abbreviations were introduced in (5.26)

$$R^i_{jk} := \ell^j R^i_{j\ kl} \tag{5.27}$$

$$P^i_{kl} := \ell^i P^i_{j\ kl} \tag{5.28}$$

The following identities result from equating equations (5.25) and (5.26)

$$R_{ijkl} + R_{jikl} + 2A_{iju} R^u_{kl} = 0 \tag{5.29}$$

$$P_{ijkl} + P_{jikl} + 2A_{iju} P^u_{kl} - 2A_{ijl|k} = 0 \tag{5.30}$$

$$A_{ijk;l} - A_{ijl;k} = A_{ijk}\ell_l - A_{ijl}\ell_k. \tag{5.31}$$

By using equation (5.29) one could derive the following relations for R^i_{jkl}.

1. $v^i g_{ip} R^p_{kl} \ = \ 0.$

2. $R^p_{ik} \ = \ R^p_{ki} - R^p_{ik}.$

3. $g_{ip} R^p_k \ = \ g_{kp} R^p_i.$

Similarly, by (5.30), the following relations hold true for P^i_{jkl}

$$g_{ip} P^p_{jkl} = -C_{ijp} P^p_{kl} - C_{jkp} P^p_{il} - C_{ikp} P^p_{jl}$$
$$- C_{ijl|k} - C_{jkl|i} - C_{ikl|j}.$$

Furthermore, by (5.30), we can easily obtain the following

1. $L_{ikl} \ = \ -g_{ip} P^p_{kl},$

2. $L_{ikl} \ = \ v^j g_{jp} P^p_{ikl},$

3. $v^j g_{jp} P^p_{kl} \ = \ L_{ikl} v^i \ = \ 0.$

5.1.3 Riemann Curvature

The Riemann curvature of a Finsler space is a family of linear transformations on tangent spaces. From the structural equations for Finsler metrics we saw that the Riemann curvature was due to the commutation operator over the sections. We now show how it could be derived via the variations of geodesics and this derivation gives us the geometric intuition. Let (M, F) be a Finsler space. Consider a geodesic $c(t)$, $a \leq t \leq b$. A C^∞ map $H : (-\epsilon, \epsilon) \times [a, b] \to M$ is called a geodesic variation of c if

1. $H(0, t) = c(t)$;

2. for each $u \in (-\epsilon, \epsilon)$, the curve $c_u(t) := H(u, t)$ is a geodesic.

For any geodesic variation, the variation field $J(t) := \frac{\partial H}{\partial u}(0, t)$ satisfies a special system of second order ordinary differential equations. The following theorem states this result.

Theorem 5.1.1. *Let (M, F) be a Finsler space. There is a family of transformations $R_v : T_xM \to T_xM$, $v \in T_xM \setminus \{0\}$, such that for any geodesic variation H of a geodesic c, the variation vector field $J(t) := \frac{\partial H}{\partial u}(0, t)$ along c satisfies the following equation*

$$D_{\dot{c}} D_{\dot{c}} J + R_{\dot{c}}(J) = 0. \tag{5.32}$$

Proof. By assumption, each $c_u(t) = H(u, t)$ is a geodesic. Thus

$$\frac{\partial^2 H^i}{\partial t^2} + 2G^i\left(H, \frac{\partial H}{\partial t}\right) \;=\; 0. \tag{5.33}$$

For simplicity, let

$$T = T^i \frac{\partial}{\partial x^i} := \frac{\partial H}{\partial t} \qquad U = U^i \frac{\partial}{\partial x^i} := \frac{\partial H}{\partial u}. \tag{5.34}$$

Equation (5.33) becomes

$$\frac{\partial T^i}{\partial t} + 2G^i(H, T) \; = \; 0 \tag{5.35}$$

Note that

$$\frac{\partial T^i}{\partial u} = \frac{\partial}{\partial u}\left(\frac{\partial H^i}{\partial t}\right) = \frac{\partial}{\partial t}\left(\frac{\partial H^i}{\partial u}\right) = \frac{\partial U^i}{\partial t} \tag{5.36}$$

Differentiating (5.35) with respect to u yields

$$\frac{\partial^2 U^i}{\partial t^2} = -2U^k \frac{\partial G^i}{\partial x^k}(H, T) - 2\frac{\partial U^j}{\partial t}\frac{\partial G^i}{\partial \dot{x}^j}(H, T) \tag{5.37}$$

Observe that

$$\frac{\partial}{\partial u}[G^i(H, T)] \;\; = \;\; U^k \frac{\partial G^i}{\partial x^k} + \frac{\partial U^j}{\partial t}\frac{\partial G^i}{\partial \dot{x}^j} \tag{5.38}$$

$$\begin{aligned}\frac{\partial}{\partial t}\left[\frac{\partial G^i}{\partial \dot{x}^j}(H, T)\right] \;\; &= \;\; T^k \frac{\partial^2 G^i}{\partial x^k \partial \dot{x}^j} + \frac{\partial T^k}{\partial t}\frac{\partial^2 G^i}{\partial \dot{x}^j \partial \dot{x}^k} \\ &= \;\; T^k \frac{\partial^2 G^i}{\partial x^k \partial \dot{x}^j} - 2G^k \frac{\partial^2 G^i}{\partial \dot{x}^j \partial \dot{x}^k}\end{aligned} \tag{5.39}$$

From the above equations (5.38) and (5.39), one obtains

$$D_T D_T U \;\; = \;\; D_T\left[\left(\frac{\partial U^i}{\partial t} + U^j \frac{\partial G^i}{\partial \dot{x}^j}\right)\frac{\partial}{\partial x^i}\right] \tag{5.40}$$

$$= \;\; -U^k \left\{2\frac{\partial G^i}{\partial x^k} - T^j \frac{\partial^2 G^i}{\partial x^j \partial \dot{x}^k} + 2G^j \frac{\partial^2 G^i}{\partial \dot{x}^j \partial \dot{x}^k} - \frac{\partial G^i}{\partial \dot{x}^j}\frac{\partial G^j}{\partial \dot{x}^k}\right\}\frac{\partial}{\partial x^i} \tag{5.41}$$

$$= \;\; -U^k R^i_k(H, T)\frac{\partial}{\partial x^i}, \tag{5.42}$$

where

$$R^i_k(x,v) \ := \ 2\frac{\partial G^i}{\partial x^k} - T^j \frac{\partial^2 G^i}{\partial x^j \partial \dot{x}^k} + 2G^j \frac{\partial^2 G^i}{\partial \dot{x}^j \partial \dot{x}^k} - \frac{\partial G^i}{\partial \dot{x}^j}\frac{\partial G^j}{\partial \dot{x}^k} \qquad (5.43)$$

For every vector $v \in T_xM \setminus \{0\}$, we define a linear transformation

$$\mathrm{R}_v = R^i_k(x,v)\frac{\partial}{\partial x^i} \otimes dx^k\Big|_x : T_xM \to T_xM. \qquad (5.44)$$

We then obtain

$$D_T D_T U + R_T U = 0.$$

Restricting the above equation to the curve c, we obtain an equation for $J(t) := U(0,t)$,

$$D_{\dot{c}} D_{\dot{c}} J + R_{\dot{c}} J = 0, \qquad (5.45)$$

which completes the proof. □

The geodesic variations give rise to a family of transformations

$$R = \{R_v : T_xM \to T_xM \mid v \in T_xM \setminus \{0\}, x \in M\}. \qquad (5.46)$$

We call it the Riemann curvature. It has the following properties

1. From equation (5.43) it follows that

$$R_v(v) = 0. \qquad (5.47)$$

2. R_v is self-adjoint with respect to g_v,

$$g_v(R_v(u), w) = g_v(u, R_v(w)), \qquad u, w \in T_xM. \tag{5.48}$$

3. From equations (5.47) and (5.48) it follows that

$$g_v(R_v(u), v) = g_v(u, R_v(v)) = 0. \tag{5.49}$$

5.1.4 Flag Curvature

The flag curvature is a geometrical invariant that generalizes the sectional curvature of Riemannian geometry. From the point of view of this chapter the flag curvature is very important as its sign tells us if two geodesics starting at a point with different initial controls diverge or converge.

Let $P \subset T_xM$ be a tangent plane. For a vector $v \in P \setminus \{0\}$, define

$$K(v, P) := \frac{g_v(R_v(u), u)}{g_v(v, v)g_v(u, u) - g_v(v, u)g_v(u, v)}, \tag{5.50}$$

$$= \frac{u^i(v^j R_{jikl} v^l)u^k}{g_v(v, v)g_v(u, u) - g_v(v, u)^2} \tag{5.51}$$

where $u \in P$ such that $P = \mathsf{span}\ \{v, u\}$. The number $K(v, P)$ is called the flag curvature of the flag (v, P) in T_xM. The geometric interpretation of flag curvature $K(v, P)$ is that along the vector $v \in T_xM \setminus \{0\}$ another vector $u := u^i \frac{\partial}{\partial x^i} \in P$ is taken such that it is transverse to the vector v. There is no loss of generality in choosing only transverse edges u that are g-orthogonal to v as any arbitrary vector w can be decomposed as $w = u + \xi\ell$, where $g(\ell, u) = 0$ and ξ is a scalar multiple. So the flag curvature $K(v, P)$ could alternately be denoted by $K(v, u)$. It is very trivial to prove that the flag curvature equation (5.51) does not change upon substituting the given direction v by the unit vectors ℓ in that direction such that $g(\ell, \ell) = 1$, i.e.,

$$K(v, u) = K(\ell, u).$$

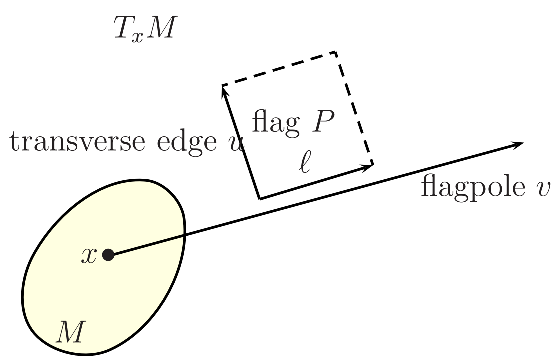

Figure 5.1: Flag on T_xM

When $n = \dim M = 2$,

$$K(v) := K(v, T_xM), \qquad v \in T_xM \setminus \{0\} \tag{5.52}$$

is a scalar function and is called the Gauss curvature. For a vector $v \in T_xM \setminus \{0\}$, there are infinitely many tangent planes $P \subset T_xM$ containing v. The flag curvature $K(P, v)$ depends on the tangent plane containing v. A Finsler metric F on a manifold M is said to be of scalar curvature $K(v)$ if $K(v, P) = K(v)$ is independent of the tangent plane P containing v for all $v \in T_xM$. This is equivalent to the following

$$R_v(u) = K(v)\{g_v(v,v)u - g_v(v,u)v\}, \qquad v, u \in T_xM \setminus \{0\}. \tag{5.53}$$

The following object is algebraically a predecessor of the flag curvature;

$$K(\ell, u, w) = \frac{u^i(\ell^j R_{jikl}\ell^l)w^k}{g_\ell(u,w) - g_\ell(\ell,u)g_\ell(\ell,w)} \tag{5.54}$$

$$= \frac{u^i R_{ik} w^k}{g_\ell(u,w) - g_\ell(\ell,u)g_\ell(\ell,w)} \tag{5.55}$$

from which the following identities follow easily:

1. $K(\ell, u, u) = K(\ell, u)$,
2. $K(\ell, u, w) = L(\ell, w, u)$,
3. $K(\ell, u, w) = \frac{1}{4}K(\ell, u + w) - \frac{1}{4}K(\ell, u - w)$.

5.2 Dispersion

5.2.1 Geodesic Variations

Let (M, F) be a Finsler space. Let $c : [a, b] \to M$ be a geodesic and $H : (-\varepsilon, \varepsilon) \times [a, b] \to M$ a geodesic variation of c, namely, each $c_s(t) := H(s, t)$ is a geodesic. Let $J(t) := \frac{\partial H}{\partial s}(0, t)$. By Lemma (5.1.1), we know that $J(t)$ satisfies the Jacobi equation

$$D_{\dot{c}}D_{\dot{c}}J + R_{\dot{c}}J = 0. \tag{5.56}$$

Conversely, for every vector field $J(t)$ along c satisfying the Jacobi equation, there is a geodesic variation H of c whose variation field is equal to $J(t)$. Thus we call a vector field $J(t)$ along c satisfying (5.56) a Jacobi field. Fix a unit vector $v \in T_xM$

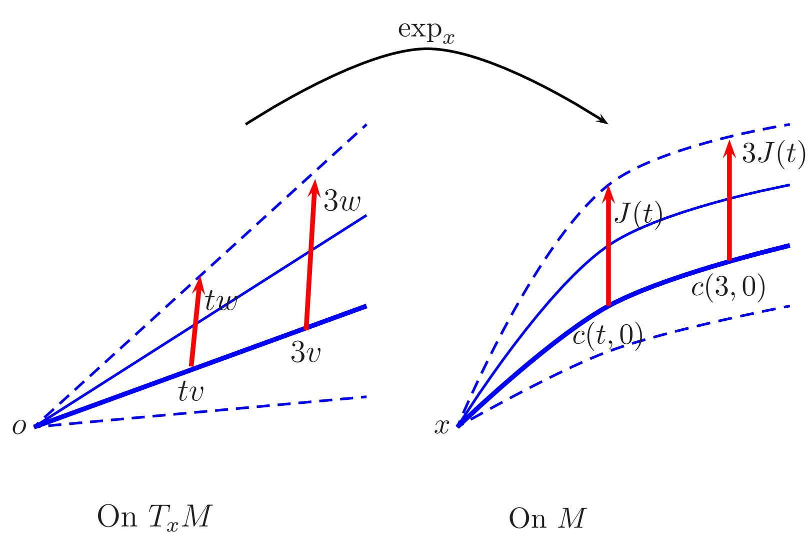

Figure 5.2: Geodesic Variations

and let $c(t) := \exp_x(tv)$, $0 \leq t \leq a$. Consider a special geodesic variation

$$H(s,t) := \exp_x[t(v + sw)], \qquad 0 \leq t < a, |s| < \varepsilon. \tag{5.57}$$

By Lemma (5.1.1),

$$J(t) := \frac{\partial H}{\partial s}(0,t) \tag{5.58}$$

is a Jacobi field along c.

Lemma 5.2.1. *The Jacobi field $J(t)$ in (5.58) is C^∞ along $c(t) = \exp_x(tv)$, $0 \leq t < a$. It satisfies the following initial conditions:*

$$J(0) = 0, \qquad D_{\dot{c}}J(0) = v. \tag{5.59}$$

5.2.2 Small time Geodesic Spread

Let $c : (-\varepsilon, \varepsilon) \to M$ be a unit speed geodesic with $\dot{c}(0) = v \in T_xM$. For a vector $w \in T_xM$, let J be the Jacobi field along c satisfying

$$J(0) = 0, \qquad D_{\dot{c}}J(0) = w. \tag{5.60}$$

Consider the following function

$$f(t) = \|J(t)\|^2 := g_{\dot{c}(t)}(J(t), J(t)), \tag{5.61}$$

which defines the length squared of the variation vector field $J(t) = \exp_{tv}(tw)$. $f(t)$ measures the rate at which the s-th geodesic is deviating from $c(t)$.

We now look at the power series expansion of the function $f(t)$ for small time t. It is to be noted that derivative the metric g along a geodesic c is given as follows

$$\frac{d}{dt}g_{\dot{c}}(V, W) = g_{\dot{c}}(D_{\dot{c}}V, W) + g_{\dot{c}}(V, D_{\dot{c}}W). \tag{5.62}$$

We use the above equation repeatedly to compute the higher-order derivatives of

$f(t)$.

$$f^{(1)}(t) = g_{\dot{c}}(D_{\dot{c}}J, J) + g_{\dot{c}}(J, D_{\dot{c}}J)$$

$$= 2g_{\dot{c}}(J, D_{\dot{c}}J) \tag{5.63}$$

$$f^{(2)}(t) = 2g_{\dot{c}}(D_{\dot{c}}J, D_{\dot{c}}J) + 2g_{\dot{c}}(J, D_{\dot{c}}D_{\dot{c}}J),$$

$$= 2g_{\dot{c}}(D_{\dot{c}}J, D_{\dot{c}}J) - 2g_{\dot{c}}(J, R(J, \dot{c})\dot{c}), \tag{5.64}$$

In the above equation we have used the Jacobi equation

$$D_{\dot{c}}D_{\dot{c}}J + R_{\dot{c}}(J) = 0,$$

where $R_{\dot{c}}(J) = R(J, \dot{c})\dot{c}$.

$$f^{(3)}(t) = 2g_{\dot{c}}(D_{\dot{c}}D_{\dot{c}}J, D_{\dot{c}}J) + 2g_{\dot{c}}(D_{\dot{c}}D_{\dot{c}}J, D_{\dot{c}}J)$$

$$+ 2g_{\dot{c}}(DDJ, DJ) + 2g_{\dot{c}}(J, D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}J)$$

$$= 6g_{\dot{c}}(D_{\dot{c}}D_{\dot{c}}J, D_{\dot{c}}J) + 2g_{\dot{c}}(J, -D_{\dot{c}}R_{\dot{c}}(J) - R_{\dot{c}}(J))$$

$$= -8g_{\dot{c}}(R_{\dot{c}}(D_{\dot{c}}J), D_{\dot{c}}J) - 2g_{\dot{c}}(R_{\dot{c}}J, J) \tag{5.65}$$

We have used the following relation

$$D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}J = -D_{\dot{c}}R_{\dot{c}}(J) - R_{\dot{c}}(J).$$

which we get by differentiating the Jacobi field equation.

$$f^{(4)}(t) = 8g_{\dot{c}}(D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}J, D_{\dot{c}}J) + 6g_{\dot{c}}(D_{\dot{c}}D_{\dot{c}}J, D_{\dot{c}}D_{\dot{c}}J) + 2g_{\dot{c}}(J, D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}J)$$

$$\begin{aligned} &= -8g_{\dot{c}}(R_{\dot{c}}(D_{\dot{c}}J), D_{\dot{c}}J) + 8g_{\dot{c}}(R_{\dot{c}}(J), R_{\dot{c}}(J)) \\ &- 12g_{\dot{c}}(\dot{R}_{\dot{c}}(J), D_{\dot{c}}J) - 2g_{\dot{c}}(\ddot{R}_{\dot{c}}(J), J). \end{aligned} \tag{5.66}$$

At the initial time, $t = 0$, we have the following data

$$J(0) = 0, \tag{5.67}$$

$$D_{\dot{c}}J(0) = J'(0) \ = \ w, \tag{5.68}$$

$$D_{\dot{c}}D_{\dot{c}}J(0) = 0, \tag{5.69}$$

$$D_{\dot{c}}D_{\dot{c}}D_{\dot{c}}J(0) = -R(w, \dot{c})\dot{c}. \tag{5.70}$$

Substituting the initial data (5.67-5.70) in the derivatives of $f(t)$ given by equations (5.63-5.66) we have

$$\begin{aligned} f(0) &= 0, \\ f^{(1)}(0) &= 0, \\ f^{(2)}(0) &= 2g_{\dot{c}}(w, w), \\ f^{(3)}(0) &= 0, \\ f^{(4)}(0) &= -8g_{\dot{c}}(R(w, \dot{c})\dot{c}, w). \end{aligned}$$

Hence the following proposition easily follows:

Proposition 5.2.2.

$$g_{\dot{c}(t)}(J,J) = g_v(w,w)t^2 - \frac{1}{3}g_v(R_v(w),w)t^4 + \mathcal{O}(t^5), \tag{5.71}$$

$$g_{\dot{c}(t)}(D_{\dot{c}}J,J) = g_v(w,w)t - \frac{2}{3}g_v(R_v(w),w)t^3 + \mathcal{O}(t^4). \tag{5.72}$$

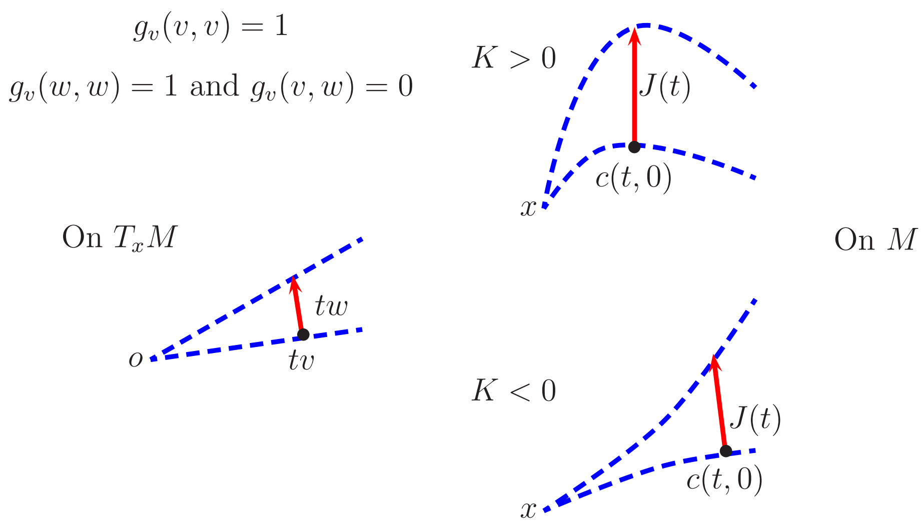

Figure 5.3: Geodesic Dispersion

From the definition of flag curvature $K(v,u)$, it is easy to see that

$$g_v(R_v(w),w) = K(v,w)\left[g_v(v,v)g_v(w,w) - g_v(v,w)^2\right]. \tag{5.73}$$

By choosing the vector v such that $g_v(v,v) = 1$ and the vector w such that $g_v(w,w) = 1$ and $g_v(v,w) = 0$, and using (5.73), equation (5.71) can be written as follows:

$$g_{\dot{c}}(J,J) = t^2 - \frac{1}{3}K(v,w)t^4 + O(t^5). \tag{5.74}$$

It can be observed from equation (5.74) that if the flag curvature at x is positive, then the geodesic trajectories emanating from x spread apart more slowly than the corresponding rays in T_xM. If the flag curvature is negative, then geodesic trajectories spread apart much faster than the corresponding rays in T_xM. Figure 5.3 depicts the spread of geodesics.

5.2.3 Neighboring Optimal Control

In order to derive control laws in presence of disturbances one needs to derive controls as a function of state, $u(x)$, which are usually called feedback controls. The exact synthesis problem in optimal control theory remains one of the unsolved problems till date. However, if one assumes that the disturbances in the system dynamics are small in magnitude and bounded then one should be able to derive those optimal controls as a function of those disturbances of state. In this chapter we deal with this problem of synthesizing controls which takes into account the geometry of the given optimal control problem.

Analogous to the literature in Riemannian geometry of taking the second variation of the energy function, we compute the second variations of the energy function of the optimal control problem on Finsler manifolds. This in turn gives us a method to approximate optimal controls with respect to a corresponding reference trajectory.

Given the corresponding optimal control problem on Finsler manifolds with a metric $F_x(\mathrm{x}')$ on the tangent space $\mathrm{x}' = (\mathrm{t}', \mathrm{x}', \mathrm{c}, \mathrm{p})$, we compute the geodesic equations as shown in the previous chapter as ,

$$\mathrm{x}''^i + \Gamma^i_{jk}(\mathrm{x}, \mathrm{x}')\mathrm{x}'^j \mathrm{x}'^k = 0, \text{ for } i = 1, \ldots, 2n + m + 1. \tag{5.75}$$

As shown in the previous chapter, we compute the optimal controls $\mathrm{c}(s)$ by solving the two-point boundary value problem by appropriate numerical techniques shown therein. The corresponding optimal state trajectory $\mathrm{y}(s)$ is called the reference trajectory connecting the initial point $\mathrm{y}(0) = \mathrm{y}_0$ and the goal $x(T) = x_T$.

Given a geodesic $x(s)$ corresponding to state conditions $x(0) = x_0, x(t_f) = x_T$, call it the reference geodesic. We now suppose a congruence of geodesics which are parameterized by two parameters $s,\ p,\ \ X^i := X^i(s, p)$, where s is the usual arc-length and p is a variation parameter such that $X^i(s, 0) = x^i(s)$, i.e., when there is no variation we retrieve our reference geodesic $x^i(s)$.

As we want to compute the corresponding optimal controls $u(s, p)$ which steers the state from an initial point $X(0, p)$, close to the initial point of the reference trajectory $X(0, 0) = x_0$, to the goal x_T with optimum cost(time), we would like to know how a small change in the parameter p of $X(s, p)$ reflects in the corresponding optimal controls $u(s, p)$.

Given the geodesic equations in the variables (y, y', c, p), we would like to find out the relations between their variations $(\delta y, \delta y', \delta c, \delta p)$ so that the they satisfy the new geodesic equations. It is very easy to show that when the covariant derivative in the Jacobi equation (5.32) that we derived earlier is expanded then we shall obtain the geodesic deviation equation that we show below. The procedure that we show here is an easier way of getting the same result.

We first obtain Taylor expansions of the Christoffel symbols of the neighboring geodesic as follows:

$$\mathrm{x}^i(s,p) = \mathrm{x}^i(s,0) + \delta\mathrm{x}^i(s) + \frac{1}{2!}\delta^2\mathrm{x}^i(s) + \frac{1}{3!}\delta^3\mathrm{x}^i(s) + \cdots \tag{5.76}$$

$$\mathrm{x}'^i(s,p) = \mathrm{x}'^i(s,0) + \delta\mathrm{x}'^i(s) + \frac{1}{2!}\delta^2\mathrm{x}'^i(s) + \frac{1}{3!}\delta^3\mathrm{x}'^i(s) + \cdots \tag{5.77}$$

$$\mathrm{x}''^i(s,p) = \mathrm{x}''^i(s,0) + \delta\mathrm{x}'^i(s) + \frac{1}{2!}\delta^2\mathrm{x}''^i(s) + \frac{1}{3!}\delta^3\ddot{x}^i(s) + \cdots \tag{5.78}$$

$$\Gamma^i_{jk}(\mathrm{x}(s,p),\mathrm{x}'(s,p)) = \Gamma^i_{jk}(\mathrm{x}(s,0),\mathrm{x}'(s,0)) + \frac{\partial\Gamma^i_{jk}}{\partial\mathrm{x}^l}\delta\mathrm{x}^l + \frac{\partial\Gamma^i_{jk}}{\partial\mathrm{x}'^l}\delta\mathrm{x}'^l + \cdots \tag{5.79}$$

By substituting the above expansions in the geodesic equation (5.75) and collecting terms of the order $o(\epsilon)$, we have the following geodesic deviation equation:

$$\frac{d^2\delta\mathrm{x}^i}{ds^2} + 2\Gamma^i_{jk}\,\frac{d\mathrm{x}^j}{ds}\,\frac{d\delta\mathrm{x}^k}{ds} + \frac{\partial\Gamma^i_{jk}}{\partial\mathrm{x}^l}\,\frac{d\mathrm{x}^j}{ds}\,\frac{d\mathrm{x}^k}{ds}\,\delta\mathrm{x}^l + \frac{\partial\Gamma^i_{jk}}{\partial\mathrm{x}^l_s}\,\frac{d\mathrm{x}^j}{ds}\,\frac{d\mathrm{x}^k}{ds}\,\frac{d\delta\mathrm{x}^l}{ds} = 0. \tag{5.80}$$

The above equation can also be written in the spray notation as follows

$$\frac{d^2\delta\mathrm{x}^i}{ds^2} + \frac{\partial G^i}{\partial\mathrm{x}^j}\delta\mathrm{x}^j + \frac{\partial G^i}{\partial\mathrm{x}'^j}\frac{d\delta\mathrm{x}^j}{ds} = 0. \tag{5.81}$$

Notice that the only difference from the Riemannian case is the last term of the

L.H.S of the above equation which accounts for a change in the Christoffel symbol due to a change in the velocity vector.

When substituted in the geodesic equations and collecting terms of the first order in δ on the both sides we get

$$\frac{d^2\delta \mathrm{x}^i}{ds^2} = \Gamma^i_{jk}\left(2\mathrm{y}'^j\frac{d\delta \mathrm{y}^k}{ds} + \frac{d\delta \mathrm{y}^j}{ds}\mathrm{c}^k + \frac{d\delta \mathrm{y}^j}{ds}\mathrm{p}^k\right) + \Gamma^i_{jk}\left(2\mathrm{p}^j\delta \mathrm{p}^k + \mathrm{c}^j\delta \mathrm{p}^k + +\mathrm{y}'^j\delta \mathrm{p}^k\right) \tag{5.82}$$

$$+ \Gamma^i_{jk}\left(2\mathrm{c}^j\delta \mathrm{c}^k + \delta \mathrm{c}^j \mathrm{p}^k + \mathrm{y}'^j\delta \mathrm{c}^k\right)$$

$$+ \left(\frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}^l}\mathrm{y}'^j\mathrm{y}'^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}^l}\mathrm{y}'^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}^l}\mathrm{y}'^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}^l}\mathrm{c}^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}^l}\mathrm{c}^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial x^l}\mathrm{p}^j\mathrm{p}^k\right)\delta \mathrm{y}^l$$

$$+ \left(\frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{y}'^j\mathrm{y}'^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{y}'^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{y}'^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{c}^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{c}^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{y}'^l}\mathrm{p}^j\mathrm{p}^k\right)\frac{d\delta \mathrm{y}^l}{ds}$$

$$+ \left(\frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{y}^j\mathrm{y}'^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{y}'^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{y}'^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{c}^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{c}^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{c}^l}\mathrm{p}^j\mathrm{p}^k\right)\delta \mathrm{c}^l$$

$$+ \left(\frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{y}'^j\mathrm{y}'^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{y}'^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{y}'^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{c}^j\mathrm{c}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{c}^j\mathrm{p}^k + \frac{\partial\Gamma^i_{jk}}{\partial \mathrm{p}^l}\mathrm{p}^j\mathrm{p}^k\right)\delta \mathrm{p}^l$$

We obtain similar geodesic deviation equations for u and λ, with the only difference being the term on the LHS, which are $\frac{d\delta u^i}{ds}$ and $\frac{d\delta\lambda^i}{ds}$. So the Geodesic deviation equations can be written in a compact form as follows

$$\frac{d^2\delta \mathrm{t}}{ds^2} = {}^{\mathrm{t}}\mathcal{D}\left(\mathrm{t}, \mathrm{y}, \frac{d\mathrm{t}}{ds}, \frac{d\mathrm{y}}{ds}, \mathrm{c}, \mathrm{p}\,;\, \delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.83}$$

$$\frac{d^2\delta \mathrm{y}}{ds^2} = {}^{\mathrm{y}}\mathcal{D}\left(\mathrm{t}, \mathrm{y}, \frac{d\mathrm{t}}{ds}, \frac{d\mathrm{y}}{ds}, \mathrm{c}, \mathrm{p}\,;\, \delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.84}$$

$$\frac{d\delta \mathrm{c}}{ds} = {}^{\mathrm{c}}\mathcal{D}\left(\mathrm{t}, \mathrm{y}, \frac{d\mathrm{t}}{ds}, \frac{d\mathrm{y}}{ds}, \mathrm{c}, \mathrm{p}\,;\, \delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.85}$$

$$\frac{d\delta \mathrm{p}}{ds} = {}^{\mathrm{p}}\mathcal{D}\left(\mathrm{t}, \mathrm{y}, \frac{d\mathrm{t}}{ds}, \frac{d\mathrm{y}}{ds}, \mathrm{c}, \mathrm{p}\,;\, \delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.86}$$

In the above equations the variables $\mathrm{t}(s), \mathrm{y}(s), \mathrm{t}'(s), \mathrm{y}'(s), \mathrm{c}(s), \mathrm{p}(s)$, are known as they are computed by solving the two-point boundary value problem for the reference geodesic. $\delta\mathrm{y}(0)$, and $\delta\mathrm{y}'(0)$ are also given at the initial point as they refer to the initial points of the neighboring geodesic we wish to compute. So that leaves us with solving for the variations in state trajectory, control trajectory and the adjoint variables trajectory from the optimal trajectories $\mathrm{y}(s), \mathrm{c}(s), \mathrm{p}(s)$ of the reference geodesic, given the initial variations in the state $\delta\mathrm{y}(0)$ and $\delta\mathrm{y}'(0)$. So the problem can be stated as follows:

Problem 5.2.3. *Given the optimal trajectories* $(\mathrm{y}(s), \mathrm{y}'(s), \mathrm{c}(s), \mathrm{p}(s))$ *of the reference geodesic, and the initial variation in the position and velocity of the state* $\delta\mathrm{y}(0) = \delta\mathrm{y}_0$, *and* $\delta\mathrm{y}'(0) = \delta\mathrm{y}'_0$, *solve the below geodesic deviation equations for the* *optimal variations* $(\delta x, \delta \dot{x}(s), \delta u(s), \delta \lambda(s))$

$$\frac{d^2\delta\mathrm{t}}{ds^2} = {}^{\mathrm{t}}\mathcal{D}\left(\delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.87}$$

$$\frac{d^2\delta\mathrm{y}}{ds^2} = {}^{\mathrm{y}}\mathcal{D}\left(\delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.88}$$

$$\frac{d\delta\mathrm{c}}{ds} = {}^{\mathrm{c}}\mathcal{D}\left(\delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.89}$$

$$\frac{d\delta\mathrm{p}}{ds} = {}^{\mathrm{p}}\mathcal{D}\left(\delta\mathrm{t}, \delta\mathrm{y}, \frac{d\delta\mathrm{t}}{ds}, \frac{d\delta\mathrm{y}}{ds}, \delta\mathrm{c}, \delta\mathrm{p}\right) \tag{5.90}$$

However we do not know the length of the optimal trajectory of the neighboring geodesic beforehand. Hence when solving geodesic deviation equations we might end up short of the goal for the neighboring geodesic, in which case it seems practical to extend the optimal trajectory of the reference geodesic and then compute the geodesic equations until we hit the goal.

Chapter 6

Numerical Computation of Finsler Geodesics

6.1 Introduction

The main objective of this chapter is to utilize the geometric objects that we have derived in the previous chapters to compute optimal controls. In this regard the Taylor expansion method is used as the numerical integrators for the geodesic equations. This leads to expressing them in terms of Riemannian curvature and the geodesic deviation. Using curvature terms in the computation of controls and errors is a novelty of this thesis.

Typical numerical methods for optimal control problems can be classified into direct and indirect methods. There exists many good surveys and books for these methods in the literature, see [14],[60],[70]. Indirect methods use the necessary conditions to compute optimal controls. Usually if the switching structure is known apriori then the indirect methods lead to a greater degree of accuracy. They focus on obtaining a solution to the classical necessary conditions for optimality which take the form of a two-point boundary value problem. One of the common variations on the indirect method is the steepest descent algorithm. In this method, the state equation is integrated forward using a guess for the control profile, and then the costate equation is integrated backward. The maximality condition is then used locally to find a steepest descent direction for u at a discrete number of points, and globally as a termination criterion. Backward integration of the costate equations may be avoided by using a direct shooting method (see [14],[15],[69]), in which initial values of the adjoint variables are guessed initially, and updated iteratively. There are some problems with this method, including a small region of convergence to the optimum, the need to formulate the costate equations, and the difficulties in choosing reasonable initial guesses for the adjoint variables. The multiple shooting method (see [40, 62]) was proposed to extend the small convergence region of the direct shooting method. Multiple shooting algorithms transform the problem into a multi-point boundary value problem for the state and adjoint variables. Homotopy methods [72] have also been proposed as a response to the small region of convergence. Other indirect techniques have been the finite-difference method [46] which is essentially a low-order collocation method, and standard collocation techniques [6].

On the other hand, direct methods are a result of discretization of the original problem and recasting as a finite-dimensional optimization problem, typically a nonlinear program (NLP), see [42]. They are used especially for tackling large-scale problems by using standard techniques from nonlinear programming. There are two general strategies within the framework of the direct method: sequential method and simultaneous strategy methods. In the sequential method often called control parametrization, the control variables are discretized over finite elements using polynomials or in fact any suitable basis functions (see [14],[15],[70]). The coefficients of the polynomials and the size of the finite elements then become decision variables in a master nonlinear program. Function evaluation is carried out by solution of an initial value problem of the original dynamic system, and gradients for a gradient-based search may be evaluated by solving either the adjoint equations or the sensitivity equations. In the simultaneous strategy both the controls and the state variables are discretized using polynomials on finite elements, and the coefficients and elements sizes become decision variables in a much larger NLP (see [75]). Unlike control parametrization, the simultaneous method does not require the solution of initial value problems at every iteration of the NLP.

There are other methods like the ε-method by Balakrishnan [7] where the dynamic constraints are entirely removed by a combination of penalty function methods and Rayleigh-Ritz type expansion methods. It has been shown that the optimal values of the penalized problem converge to the optimal value of the relaxed problem.

Techniques such as gradient methods in function spaces using penalty function methods and feasible directions methods can be found in [55],[56],[60]. Iterative methods which are essentially based on solutions of the Riccati differential equations appearing in linear quadratic optimal control problems are also used in computing optimal controls, see [41], [37]. Such methods solve the original nonlinear problem through a sequence of linear quadratic optimal control problems which are suitably formulated as the accessory optimal control problems to the original one. This method exhibits fast convergence properties. However for large scale problems it is avoided due to derivative computations.

Finally, the Hamilton-Jacobi-Bellman equation which is derived from Bellman's principal of optimality has its own set of techniques. However this is a partial differential equation and in practice it is very difficult to solve except in certain fortuitous cases, see [10], [9]. There has been a lot of research in computing optimal controls in the aerospace, chemical processes, and recently finance communities.

6.2 Initial Value Problem: Geodesics

We consider Taylor series type numerical integrators for the initial value problems. This gives us an entirely symbolic method to integrate the geodesics and furthermore we can use all the geometric quantities that we have obtained before in the series expansion. All the terms in the series in the Taylor expansion can be written in terms of the curvature operator. We first give an outline of Taylor polynomial solution to ordinary differential equations.

6.2.1 Taylor Methods

Let E be an open set in $\mathbb{R}^n$, $x_0 \in E$ and $f : E \to \mathbb{R}^n$ be analytic. Consider the initial value problem

$$\dot{x} = f(x), \qquad x(0) = x_0, \tag{6.1}$$

and let J be the maximal interval of existence of the solution. Then the solution $x(t)$ is analytic on J and can be expanded in the power series

$$x(t) = \sum_{k=0}^{\infty} x_k t^k, \qquad t \in (-R, R) \subset J \tag{6.2}$$

where $R > 0$ is the radius of convergence of the series and

$$x_k = \frac{d^k}{dt^k} x(t)\Big|_{t=0}, \qquad k = 0, 1, \ldots \tag{6.3}$$

are the Taylor coefficients of the solution curve. It is well-known that the derivative $\dot{x}(t)$ is analytic as well and

$$\dot{x}(t) = \sum_{k=0}^{\infty} (k+1) x_{k+1} t^k, \qquad -R < t < R. \tag{6.4}$$

Let

$$f(x(t)) = \sum_{k=0}^{\infty} f_k t^k, \qquad -R < t < R \tag{6.5}$$

be the power series expansion of f along the curve $x(t)$. Then comparing the coefficients in the equation $\dot{x}(t) = f(x(t))$ gives the following recurrence relation for the Taylor coefficients $\{x_k\}$:

$$x_{k+1} = \frac{1}{k+1} f_k, \qquad k = 0, 1, \ldots. \tag{6.6}$$

Taylor Series of degree 1

Consider an interval $I_n = [t_n, t_{n+1}]$ with a uniform stepsize $t_{n+1} - t_n = h$. Then a Taylor series of first degree expanded about t_n for (6.1) is given by the equation

$$x(t_{n+1}) = x(t_n) + hf(t_n, x_n) + \frac{h^2}{2} f'(\xi_n, x(\xi_n)), \qquad t_n < \xi_n < t_{n+1} \tag{6.7}$$

from which one obtains the explicit(forward) Euler's method

$$x_{n+1} = x_n + hf(t_n, x_n) \tag{6.8}$$

The error may also be expressed by the expansion

$$e_{n+1} := x(t_{n+1}) - x_{n+1} = \frac{h^2}{2} f_n' + \frac{h^3}{3} f_n'' + \frac{h^4}{4} f_n''' + \cdots \tag{6.9}$$

where $f_n' := f'(t_n, x_n)$, and similarly for f_n'' and f_n'''.

If the series is expanded about t_{n+1}, one obtains the backward Euler method

$$x_{n+1} = x_n + hf(t_{n+1}, x_{n+1}), \tag{6.10}$$

where x_{n+1} must be obtained implicitly. By expanding $f(t_{n+1}, x_{n+1})$ in a Taylor series about (t_n, x_n), one derives from (6.10) the discretization error

$$\begin{aligned} e_{n+1} \;&:=\; x(t_{n+1}) - x_{n+1} = -\frac{h^2}{2} f_n' - \frac{h^3}{3}\left(2f_n'' + 3\frac{\partial f_n}{\partial x} f_n'\right) \\ &-\frac{h^4}{4}\left(3f_n''' + 8\frac{\partial f_n}{\partial x} f_n'' + 12\frac{\partial f_n'}{\partial x} f_n'\right) + O(h^5) \end{aligned} \tag{6.11}$$

Both the forward and backward Euler methods provide a linear approximation to x on $I_n := [t_n, t_{n+1}]$. This approximation can be written as

$$\tilde{x}_n = x_{n+\mu} + (t - t_{n+\mu}) f(t_{n+\mu}, x_{n+\mu}), \qquad t_n \leq t \leq t_{n+1}, \tag{6.12}$$

where $\mu \in \{0, 1\}$, respectively. If we define $t_{n+\mu} := t + \mu h$, where $0 \leq \mu \leq 1$, then equation (6.12) becomes a generalized approximation to x on I_n. In this case the approximation will pass through the intermediate point $x_{n+\mu}$ which is unknown. By observing that $\tilde{x}_n$ must equal x_n when $t = t_n$, one can see that $x_{n+\mu}$ must be determined from

$$x_{n+\mu} = x_n + \mu h f(t_{n+\mu}, x_{n+\mu}). \tag{6.13}$$

Note that this is just the backward Euler method for the partial step from t_n to $t_{n+\mu}$. Once $x_{n+\mu}$ is determined from (6.13), one sets $t = t_{n+1}$, $\tilde{x}_n = x_{n+1}$ in (6.12) to obtain

$$x_{n+1} = x_{n+\mu} + (1 - \mu) h f(t_{n+\mu}, x_{n+\mu}). \tag{6.14}$$

One observes that this is just the forward Euler method for the partial step from $t_{n+\mu}$ to t_{n+1}. Thus, two steps are required to advance the solution to t_{n+1}, i.e., a backward Euler intermediate step followed by a forward Euler final step. To determine an optimal expansion point $t_{n+\mu}$, one obtains the truncation error for (6.13). Since (6.13) is the backward Euler method with stepsize μh, one uses (6.11) to obtain an expansion for $x_{n+\mu}$, which may then be substituted into (6.14). One gets from (6.11),

$$x_{n+\mu} = x_n + \mu h f_n + (\mu h)^2 f_n' + \frac{(\mu h)^2}{2}\left(f_n'' + \frac{\partial f_n}{\partial x} f_n'\right) + O(h^4). \tag{6.15}$$

Upon expanding $f(t_{n+\mu}, x_{n+\mu})$ about (t_n, x_n), and evaluating powers of $(x_{n+\mu} - x_n)$ by way of (6.15), one obtains from (6.14)

$$x_{n+1} = x_n + h f_n + \mu h^2 f_n' + \mu^2 \frac{h^3}{2}\left(f_n'' + \frac{\partial f_n}{\partial x} f_n'\right) + O(h^4). \tag{6.16}$$

The error expansion for (6.13-6.14) is then given by

$$e_{n+1} := x(t_{n+1}) - x_{n+1} = (1-2\mu)\frac{h^2}{2} f_n' + \frac{h^3}{6}\left[(1-3\mu^2) f_n'' - 3\mu^2 \frac{\partial f_n}{\partial x} f_n'\right] \tag{6.17}$$

Several other marching schemes can be obtained from the generalized approximation (6.12).

Taylor Series of degree 2

Taylor series of second degree expanded about the point $t_{n+\mu}$ approximates x on the interval I_n by

$$x_n = x_{n+\mu} + (t - t_{n+\mu}) f(t_{n+\mu}, x_{n+\mu}) + \frac{(t - t_{n+\mu})^2}{2} f'(t_{n+\mu}, x_{n+\mu}). \tag{6.18}$$

As in the previous section, the expansion coefficient μ defines a family of marching schemes. Choosing $\mu = 0$ and setting $t = t_{n+1}$, $z_n = x_{n+1}$, one obtains the explicit

Taylor series method of order 2

$$x_{n+1} = x_n + hf(t_n, x_n) + \frac{h^2}{2} f'(t_n, x_n) \tag{6.19}$$

with local error

$$e_{n+1} = \frac{h^3}{6} f_n'' + \frac{h^4}{24} f_n''' + \cdots . \tag{6.20}$$

Similarly, choosing $\mu = 1$ and setting $t = t_n$, $z_n = x_n$, one obtains the implicit method

$$x_{n+1} = x_n + hf(t_{n+1}, x_{n+1}) - \frac{h^2}{2} f'(t_{n+1}, x_{n+1}) \tag{6.21}$$

with local error

$$e_{n+1} = \frac{h^3}{6} f_n'' + \frac{h^4}{24} (3f_n''' + 4\frac{\partial f_n}{\partial x} f_n'') + \cdots . \tag{6.22}$$

When $0 < \mu < 1$, (6.18) leads to a two-half step, implicit/explicit marching scheme. The implicit half step is given by

$$x_{n+\mu} = x_n + \mu h f(t_{n+\mu}, x_{n+\mu}) - \frac{(\mu h)^2}{2} f'(t_{n+\mu}, x_{n+\mu}), \tag{6.23}$$

and the explicit half step is given by

$$x_{n+1} = x_{n+\mu} + (1-\mu) h f(t_{n+\mu}, x_{n+\mu}) + \frac{[(1-\mu)h]^2}{2} f'(t_{n+\mu}, x_{n+\mu}). \tag{6.24}$$

The error is given by

$$e_{n+1} = (1 - 3\mu + 3\mu^2) \frac{h^3}{6} f_n'' + \frac{h^4}{24} \left[(1 - 6\mu^2 + 8\mu^3) f_n''' + 4\mu^3 \frac{\partial f_n}{\partial x} f_n'' \right] \tag{6.25}$$

6.2.2 Taylor Series Integration of Geodesics

As we have seen that the geodesic equation is given by

$$\frac{d^2x^i}{ds^2} + 2G^i\left(x, \frac{dx}{ds}\right) \;=\; 0, \tag{6.26}$$

where the G^i are the geodesic coefficients which are positively homogeneous of degree 2 in $\frac{dx}{ds}$. Now let us consider a geodesic C through a fixed point O with coordinates a_0^i, its tangent at O being denoted by Consider the initial value problem

$$x_s^i \;=\; y^i \tag{6.27}$$

$$y_s^i \;=\; -\Gamma^i_{jk}y^jy^k \tag{6.28}$$

$$x^i(0) \;=\; x_0^i \tag{6.29}$$

$$y^i(0) \;=\; y_0^i \tag{6.30}$$

The Taylor's series series expansion of a curve C at a point p given by $x^i = x^i(0)$ with initial direction $x'^i = x'^i(0)$ is given by

$$x^i(s) = a_0^i + a_1^i s + a_2^i(0)\frac{s^2}{2!} + a_3^i\frac{s^3}{3!} + \ldots \tag{6.31}$$

where $a_0^i = x^i(0)$ and $a_1^i = x'^i(0)$. One can compute the coefficients $\{a_j^i\},\ j = 2, \ldots$ by taking higher order derivatives of the geodesic equation (6.26) and obtaining them as a function of (a_0^i, a_1^i). The coefficients a_j^i upto $j = 4$ are then given as follows:

$$a_0^i = x^i(0) \tag{6.32}$$

$$a_1^i = x_s^i(0) \tag{6.33}$$

$$a_2^i = -2\, G^i(a_0, a_1) \tag{6.34}$$

$$a_3^i = -2\, a_1^l\, \frac{\partial G^i}{\partial x^l}(a_0, a_1) + 2\, G^l(a_0, a_1)\, \frac{\partial G^i}{\partial \dot{x}^l}(a_0, a_1) \tag{6.35}$$

$$\begin{aligned} a_4^i &= -2\frac{\partial G^i}{\partial x^l} a_1^l a_1^m + 4\frac{\partial^2 G^i}{\partial \dot{x}^m \partial x^l} a_1^l G^m - 8\frac{\partial^2 G^i}{\partial \dot{x}^m \partial \dot{x}^l} G^l G^m \\ &\quad +2\frac{\partial G^i}{\partial \dot{x}^l}\frac{\partial G^i}{\partial x^l} a_1^l - 4\frac{\partial G^i}{\partial \dot{x}^l}\frac{\partial G^i}{\partial \dot{x}^m} G^m + 4\frac{\partial G^i}{\partial x^l} G^l \\ &= R_k^i G^m - 2\frac{\partial G^i}{\partial x^l} a_1^l a_1^m \end{aligned} \tag{6.36}$$

When we substitute the above derivatives in Taylor series (6.31), we get

$$\begin{aligned} x^i(h) &= a_0^i + a_1^i h - 2\, G^i \frac{h^2}{2!} + \left(-2\, a_1^l\, \frac{\partial G^i}{\partial x^l} + 2\, G^l\, \frac{\partial G^i}{\partial \dot{x}^l} \right) \frac{h^3}{3!} \\ &\quad +4\left(R_k^i G^m - 2\frac{\partial G^i}{\partial x^l} a_1^l a_1^m \right) \frac{h^4}{4!} + O(h^4) \end{aligned} \tag{6.37}$$

In the optimal control setting, given an initial value for the adjoint variable λ and the position x_0, we derive the control u by solving the maximality condition in the Pontryagin's Maximum Principle. We then use this control to calculate the initial direction $\dot{x} = f(x, u)$ and the reparametrization of time with respect to arc-length s by computing $t'(s) = 1/L(x, u)$. We then recompute our initial values of velocity, controls, and adjoint variables as a function of arc-length. This procedure is outlined in the following algorithm.

Algorithm 1 GeodesicIVP

Require: $x(0)$ and $\lambda(0)$

$\lambda(0) \leftarrow \lambda_0$

$t'(n) \leftarrow 1/L(x(n), u(n))$

$\frac{dx}{ds}(0) \leftarrow f(x(0), u(0))$

Compute $u(0) \leftarrow \arg\min_u H(x(0), u, \lambda(0))$

$a_0(0) \leftarrow (x(0), 0, 0)$

$a_1(0) \leftarrow (\dot{x}(0), u(0), \lambda(0))$

while $n \leq N$ **do**

  while $i \leq 2n + m + 1$ **do**

    $a_0^i(n+1) = a_0^i(n) + a_1^i(n)h - 2\,G^i \frac{h^2}{2!} + \left(-2\,a_1^l(n)\,\frac{\partial G^i}{\partial x^l}(n) + 2\,G^l\,\frac{\partial G^i}{\partial \dot{x}^l}(n) \right) \frac{h^3}{3!}$

    $a_1^i(n+1) = a_1^i(n) - 4\,G^i \frac{h}{2!} + 3\left(-2\,a_1^l(n)\,\frac{\partial G^i}{\partial x^l}(n) + 2\,G^l\,\frac{\partial G^i}{\partial \dot{x}^l}(n) \right) \frac{h^2}{3!}$

    $e^i(n+1) = 4\left(R_k^i(n) G^m(n) - 2\frac{\partial G^i}{\partial x^l}(n) a_1^l(n) a_1^m(n) \right) \frac{h^4}{4!}$

    $i \leftarrow i + 1$

  end while

  $n \leftarrow n + 1$

end while

Notice that in the Taylor series expansion we need to calculate the derivatives of inverses of the metric g_{ij} which appears in the geodesic equation

$$\mathsf{g}_{ij}\, x^j_{ss} = G_i \tag{6.38}$$

Let us write the above geodesic equation in the following matrix notation

$$\mathsf{g}\, x_{ss} = G.$$

We now derive the derivatives of the state x with respect to arc-length s as

$$\alpha = \mathsf{g}^{-1}G \tag{6.39}$$

$$\begin{aligned} \alpha' &= -\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}\, G + \mathsf{g}^{-1}G' \\ &= -X_1G + \mathsf{g}^{-1}G' \end{aligned} \tag{6.40}$$

$$\begin{aligned} \alpha'' &= -[\mathsf{g}^{-1}\mathsf{g}''\mathsf{g}^{-1} - 2(\mathsf{g}^{-1}\mathsf{g}')\mathsf{g}^{-1}(\mathsf{g}'\mathsf{g}^{-1})]G - 2\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}G' + \mathsf{g}^{-1}G'' \\ &= -[X_2 - 2X_1\mathsf{g}X_1]G - 2X_1G' + \mathsf{g}^{-1}G'' \end{aligned} \tag{6.41}$$

$$\begin{aligned} \alpha''' &= -[\mathsf{g}^{-1}\mathsf{g}'''\mathsf{g}^{-1} - 3[(\mathsf{g}^{-1}\mathsf{g}'')\mathsf{g}^{-1}(\mathsf{g}'\mathsf{g}^{-1}) + (\mathsf{g}^{-1}\mathsf{g}'')\mathsf{g}^{-1}(\mathsf{g}''\mathsf{g}^{-1})] + 6\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}]G \\ &\quad +3[2\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1} - \mathsf{g}^{-1}\mathsf{g}''\mathsf{g}^{-1}]G' - 3\mathsf{g}^{-1}\mathsf{g}'\mathsf{g}^{-1}\, G'' + \mathsf{g}^{-1}G''' \qquad (6.42) \\ &= -[X_3 - 3[X_2\mathsf{g}X_1 + X_1\mathsf{g}X_2] + 6X_1\mathsf{g}X_1\mathsf{g}X_1]G + 3[2X_1\mathsf{g}X_1 - X_2]G' - 3X_1G'' + \mathsf{g}^{-1}G''' \end{aligned}$$

where

$$X_1 = \mathsf{g}^{-1}\mathsf{g}^{'}\mathsf{g}^{-1}$$
$$X_2 = \mathsf{g}^{-1}\mathsf{g}^{''}\mathsf{g}^{-1} \tag{6.43}$$
$$X_3 = \mathsf{g}^{-1}\mathsf{g}^{'''}\mathsf{g}^{-1}$$

In the above computations, a lot of matrix multiplications are involved. In order to avoid these computations we first compute the derivatives without inversion and then derive the following recurrence relations

$$\mathsf{g}\,\alpha = G \tag{6.44}$$
$$\mathsf{g}\,\alpha^{'} = -\mathsf{g}^{'}\,\alpha + G^{'} \tag{6.45}$$
$$\mathsf{g}\,\alpha^{''} = -\mathsf{g}^{''}\,\alpha - 2\mathsf{g}^{'}\,\alpha^{'} + G^{''} \tag{6.46}$$
$$\mathsf{g}\,\alpha^{'''} = -\mathsf{g}^{'''}\,\alpha - 3\mathsf{g}^{''}\,\alpha^{'} - 3\mathsf{g}^{'}\,\alpha^{''} + G^{'''} \tag{6.47}$$
$$\mathsf{g}\,\alpha^{''''} = -\mathsf{g}^{''''}\,\alpha - 4\mathsf{g}^{'''}\,\alpha^{'} - 6\mathsf{g}^{''}\,\alpha^{''} - 4\mathsf{g}^{'}\,\alpha^{'''} + G^{''''} \tag{6.48}$$
$$\tag{6.49}$$

This set of computations involves a the same number of derivatives and comparatively less number of matrix multiplications. The higher order derivatives of the metric g are given as follows

$$\mathsf{g} = [F^2]_{x'^i x'^j}$$

$$\mathsf{g}' = \frac{\partial g_{ij}}{\partial x^k} a_1^k + \frac{\partial g_{ij}}{\partial v^k} a_2^k$$

$$\mathsf{g}\,\mathsf{g}' = \mathsf{g}\,\frac{\partial g_{ij}}{\partial x^k} a_1^k + \frac{\partial g_{ij}}{\partial v^k} G_k$$

$$\mathsf{g}\,\mathsf{g}'' = \frac{\partial^2 g_{ij}}{\partial x^k \partial x^l} a_1^k a_1^l + 2\frac{\partial^2 g_{ij}}{\partial v^k \partial x^l} a_2^k a_1^l + \frac{\partial^2 g_{ij}}{\partial v^k \partial v^l} G_k a_2^l$$
$$+ \frac{\partial g_{ij}}{\partial v^k}\frac{\partial G_k}{\partial x^l} a_1^l + \frac{\partial g_{ij}}{\partial v^k}\frac{\partial G_k}{\partial v^l} a_2^l$$

One could now recompute the coefficients a_j^i without the inversion of the metric:

$$g_{il}\, a_2^l = -2\, G_i(a_0, a_1) \tag{6.50}$$

$$g_{il}\, a_3^l = -g_{il}\frac{\partial g_{jk}}{\partial x^m} a_1^m a_2^l - \frac{\partial g_{jk}}{\partial \dot{x}^m} G_m$$
$$-2\, a_1^l\, \frac{\partial G_i}{\partial x^l}(a_0, a_1) + 2\, G_l(a_0, a_1)\, \frac{\partial G_i}{\partial \dot{x}^l}(a_0, a_1) \tag{6.51}$$

6.3 Boundary-Value Problem

In the boundary-value problems one seeks a solution $y(x)$ of a system of n-ODEs,

$$y' = f(x, y),$$

where

$$y = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T, \text{ and } \qquad f(x,y) = \begin{bmatrix} f_1(x,y) & \dots & f_n(x,y) \end{bmatrix}^T$$

satisfying a boundary condition of the form

$$Ay(a) + By(b) = c,$$

where $a \neq b$ are given numbers, A, B square matrices of order n, and c a vector in $\mathbb{R}^n$.

6.3.1 Simple Shooting Method

In this section we show how to implement the simple shooting method for the following boundary-value problem

$$x'' = G(x, x'), \qquad x(a) = \alpha, \ \ y(b) = \beta, \tag{6.52}$$

with separated boundary conditions. The initial-value problem

$$x'' = G(x, x'), \qquad x(a) = \alpha, \ \ x'(a) = q \tag{6.53}$$

in general has a uniquely determined solution $x(s) \equiv x(s; q)$ which of course depends on the choice of the initial value for q for $x'(a)$.

To solve the boundary value problem (6.52), we must determine $q = \tilde{q}$ so as to satisfy the second boundary condition, $y(b) = y(b; \tilde{p}) = \beta$. In other words, one has to find a zero $\tilde{p}$ of the function $F(q) := x(b; q) - \beta$.

Since $x(b; q)$, and hence $F(q)$, are in general continuously differentiable functions of q, for determining a zero $\tilde{q}$ of $F(q)$, one can use the Newton's method. Starting with an initial approximation $q^{(0)}$, one then has to iteratively compute values $q^{(i)}$ according to the prescription

$$q^{(i+1)} = q^{(i)} - (\nabla F(q^{(i)}))^{-1} \cdot F(q^{(i)}) \tag{6.54}$$

In each iteration step, therefore, one has to compute $F(q^{(i)})$, the Jacobian matrix

$$\nabla F(q^{(i)}) = \left[\frac{\partial F_j}{\partial q_{(k)}} \right]_{q=q^{(i)}},$$

and the solution $d^i := q^{(i)} - q^{(i+1)}$ of the linear system of equations

$$\nabla F(q^{(i)}) \cdot d^i = F(q^{(i)}).$$

For the computation of $F(q^{(i)}) = r(q^{(i)}, x(b; q^{(i)}))$ one must determine $x(b; q^{(i)})$, i.e., solve the initial value problem (6.53) for $q = q^{(i)}$.

In view of the the mere local convergence of the (approximate) Newton's method, it will in general diverge unless the starting vector $q^{(0)}$ is already sufficiently close to a solution $\tilde{q}$ of $F(q) = 0$. Hence, this method is not very useful in practice, and for this reason, one replaces the Newton's method by the modified Newton method, which usually converges even for starting vectors that are not particularly good, of course if at all the BVP is solvable.

Algorithm 2 GeodesicBVP

Require: x_0, Z^1

1: $i \Leftarrow 1$
2: **while** $|Z^{i+1} - Z^i| \leq \varepsilon$ **do**
3: Integrate the IVPs to obtain x and z
4: GeodesicIVP(a_0, a_1) $x''(s) + G(x, x') = 0$ subject to $a_0 = x(0)$, $a_1 = Z^i$.
5: GeodesicIVP$(0, 1)$ $z''(s) + \frac{\partial G}{\partial x} z + \frac{\partial G}{\partial x'} z' = 0$ subject to $z(0) = 0$ and $z'(0) = 1$.
6: Compute $\Phi(Z^i) = x(1, Z^i)$
7: Solve the linear system $J(Z^i)S^i = \Phi(Z^i)$ where $J(Z^i) = \frac{\partial x(1, Z^i)}{\partial s} = z(x(0), x'(0), 1)$
8: Determine $Z^{i+1} = Z^i + S^i$
9: **end while**

6.3.2 Multiple Shooting

In a multiple shooting method, the values

$$\tilde{q}_k = x(s_k), \qquad k = 1, 2, \ldots, m,$$

of the exact solution $y(x)$ of a boundary problem

$$x'' = G(x, x'), \qquad r(x(a), x(b)) = 0,$$

at several points

$$a = s_1 < s_2 < \cdots < s_m = b$$

are computed simultaneously by iteration.

Let $x(s; s_k, q_k)$ be the solution of the initial value problem

$$x'' = G(x, x'), \qquad x(s_k) = q_k.$$

The problem now consists of determining the vectors q_k, $k = 1, 2, \ldots, m$, in such a way that the function

$$x(s) \;:=\; x(s; s_k, q_k) \quad \text{for} \;\; x \in [s_k, s_{k+1}),\; k = 1, 2, \ldots, m-1,$$

$$x(b) \;:=\; q_m,$$

pieced together by the $x(s; s_k, q_k)$, is continuous, and thus a solution of the differential equation $x'' = G(x, x')$, and in addition satisfies the boundary conditions $r(x(a), x(b)) = 0$.

This yields the following nm conditions:

$$\begin{aligned} x(s_{k+1}; s_k, q_k) &= q_{k+1},\; k = 1, 2, \ldots, m-1, \\ r(q_1, q_m) &= 0. \end{aligned}$$

These system of equations could be written in the vector form as

$$F(p) := \begin{bmatrix} F_1(q_1, q_2) \\ \vdots \\ F_{m-1}(q_{m-1}, q_m) \\ F_m(q_1, q_m) \end{bmatrix} := \begin{bmatrix} x(s_2; s_1, q_1) - q_2 \\ \vdots \\ x(s_m; s_{m-1}, q_{m-1}) - q_m \\ r(q_1, q_m) \end{bmatrix} = 0$$

in the unknowns

$$p = \begin{bmatrix} q_1 \\ \vdots \\ q_m \end{bmatrix}.$$

The above optimization problem can be solved iteratively using Newton's method,

$$q^{i+1} = q^i - DF(q^i)^{-1}F(q^i), \qquad i = 0, 1, \ldots.$$

In each step of the method one must compute $F(q)$ and $DF(q)$ for $q = q^i$. The Jacobian matrix $DF(q)$ has the form

$$DF(q) = \begin{bmatrix} G_1 & -I & 0 & & 0 \\ 0 & G_2 & -I & & \\ & & \ddots & & 0 \\ 0 & & & G_{m-1} & -I \\ A & 0 & & 0 & B \end{bmatrix},$$

where the $n \times n$ matrices A, B, G_k, $k = 1, \ldots, m-1$, in turn are Jacobian matrices,

$$\begin{aligned} G_k &:= D_{q_k} F_k(q) &\equiv D_{q_k} x(s_{k+1}; s_k, q_k), \\ B &:= D_{q_m} F_m(q) &\equiv D_{q_m} r(q_1, q_m), \\ A &:= D_{q_1} F_m(s) &\equiv D_{q_1} r(q_1, q_m). \end{aligned} \tag{6.55}$$

One thus finds

$$\begin{aligned} \triangle q_2 &= G_1 \triangle q_1 + F_1, \\ &\vdots \\ \triangle p_m &= G_{m-1} G_{m-2} \cdots G_1 \triangle q_1 + \sum_{j=1}^{m-1} \left(\prod_{l=j+1}^{m-1} G_l \right) F_j, \end{aligned}$$

and from this finally, by means of the last equation,

$$(A + BG_{m-1} G_{m-2} \cdots G_1) \triangle p_1 = w,$$

where

$$w = -(F_m + BF_{m-1} + BG_{m-1} F_{m-2} + \cdots + BG_{m-1} G_{m-2} \cdots G_1 F_1).$$

This is a system of linear equations for the unknown vector $\triangle q_1$, which can be solved by means of Gaussian elimination. Once $\triangle q_1$ is determined, one obtains $\triangle q_2,\ \triangle q_3, \ldots, \triangle q_m$ successively from.

The memory required by this algorithm is essentially $m \cdot n^2$. The computational cost is dominated by the accumulation matrix E as an $(M-1)$-fold product of

(m, n)-matrices. Together with decomposition of E this results in a cost of $O(m \cdot n^3)$ operations.

In the case of $m = 3$ which is what we shall be interested as we would also like to compute the midpoint of the obtained geodesic, we have

$$
\begin{aligned}
(A + BG_2G_1)\triangle q_1 &= -r - B(F_2 - G_2F_1) \\
\triangle q_2 &= G_1 \triangle q_1 + F_1 \\
\triangle q_3 &= G_2 \triangle q_2 + F_2
\end{aligned}
$$

We shall now use the Taylor integrators for IVPs and the shooting method for BVPs to develop a method to compute optimal controls.

6.4 Computing Optimal Controls

So far we have developed the tools for solving initial value and boundary value problems. We now use these tools to compute optimal controls. In general computing global optimal controls is a nontrivial task. Their existence itself is hard to answer. We therefore settle for small time geodesics and piece them together if global ones are at all possible. The idea is to choose(guess) a control function $u(s)$ which takes us to the goal starting from an initial point x. So the job is: given an initial guess of the control function $u^{(0)}(s)$ from the set of admissible control functions $\mathcal{U}_{[0,T]}$, we need to find a sequence of control functions $u^{(j)}(s)$ which converge to the optimal control function $u^*(s)$, such that

$$|V^{i+1} - V^i| \leq \varepsilon, \tag{6.56}$$

where

$$V^i := \int_0^T L(x(t), u^{(i)}(t))dt. \tag{6.57}$$

Our algorithm begins with a subdivision of the total cost V^0 of the initial path into N intervals with nodes

$$0 < s_1 < s_2 < \cdots s_{N-1} < s_N = V^0.$$

Let the state and velocity take values

$$(x_j^{(i)}, x'^{(i)}_j)$$

at j-th node and i-th iteration.

The algorithm first finds a geodesic between the state nodes $x_1^{(0)}$ and $x_3^{(0)}$ which could be done by the simple shooting method. The next step involves transporting the midpoint $x_2^{(0)}$ of the initial trajectory to the new geodesic and rename it $x_2^{(1)}$. This process is continued by obtaining the geodesic between nodes $x_2^{(1)}$ and $x_4^{(0)}$ and choosing the midpoint as $x_3^{(1)}$. In this way we get a new trajectory $x^{(i)}(s)$ after each iteration. This technique is illustrated in the following figures.

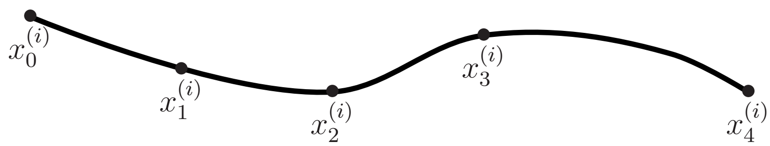

Figure 6.1: Incremental Geodesic Algorithm: a

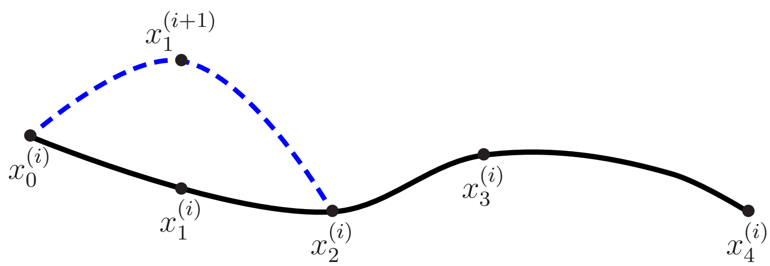

Figure 6.2: Incremental Geodesic Algorithm: b

It should be noted that trajectories starting from iteration 2 are local geodesics. Due to this fact we are in a situation where we can estimate the geodesic spread for small time as the geodesics of each iteration generate a Jacobi field. Recalling from the previous chapter that the spread given by $g_{\dot{c}}(J(t), J(t))$ can be estimated for small times. So curvature operators give us a good estimate of what initial directions one should choose based on small time geodesic spread. This method is outlined in the following algorithm.

Algorithm 3 OptControls

Require: u

$j \leftarrow 0$

$n \leftarrow 1$

while $|V^{n+1} - V^{n}| \leq \varepsilon$ and $|u^{n+1} - u^{n}| \leq \epsilon$ **do**

  while $j \leq N - 1$ **do**

    $x(s) \leftarrow GeodesicBVP(x_{j}^{(n+1)}, x_{j+2}^{(n)})$

    $x_{j+1}^{(n+1)} \leftarrow MidPoint(x(s))$

    $j \leftarrow j + 1$

  end while

  $n \leftarrow n + 1$

end while

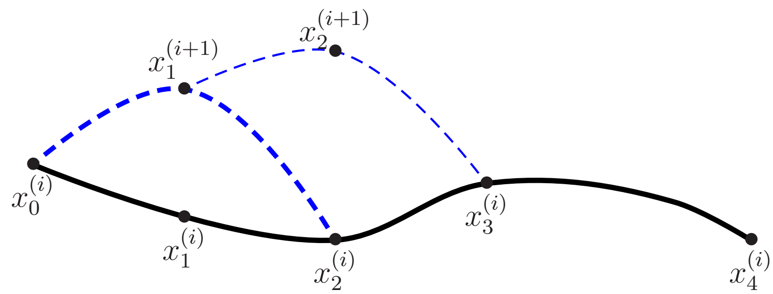

Figure 6.3: Incremental Geodesic Algorithm: c

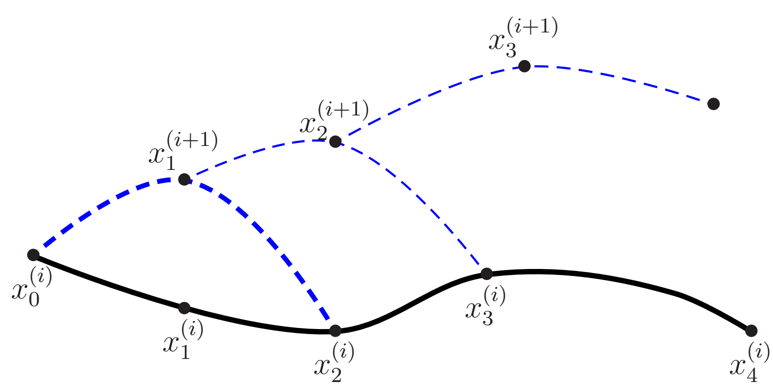

Figure 6.4: Incremental Geodesic Algorithm: d

Chapter 7

Future Research Plans

Symmetries and Conservation Laws of Optimal Control Geodesics

The Norwegian mathematician Sophus Lie pioneered the study of continuous Lie transformation groups that leave systems of differential equations invariant. As a result of Lie's work, diverse and ad hoc integration methods for solving special classes of differential equations came under a common conceptual umbrella. The classical Lie symmetry group of a system of differential equations is a local group of point transformations, meaning diffeomorphisms on the space of independent and dependent variables, which map solutions of the system into other solutions.

Conservation laws are an extension of lie symmetries in the sense that we not only consider point transformations but also transformations involving higher order derivatives. These generalized transformations are known as Lie-Backlund transformations. Emmy Noether was the first mathematician to study conservation laws, however, she only dealt with transformations upto first order derivatives. The symmetries and conservation laws involving higher order transformations are known as generalized symmetries and conservation laws. We also investigate conservation laws derived from Killing equations which are derived from invariance of the Lie derivative of covariant 2 metric tensor with respect to the infinitesimal transformations. Here the Lie derivative involves tensors which are also a function of vectors, i.e., $g_{ij}(x; x')$.

Having derived the Finsler geodesics of optimal control problems, we intend to investigate the Lie symmetries of these geodesic equations. Usually these symmetries are used to find closed form similarity solutions or reduce the differential equations to lower-order ones or be completely integrated via group theoretic techniques. We however, are not of the opinion that these geodesics can be completely solved in closed form but would like to investigate how these symmetries and conservation laws could be useful in the numerical computation of optimal feedback controls.

Infinite Dimensional Linear Programming

Optimal control problems can be reformulated as infinite dimensional linear programming problems. J.E. Rubio [63] has studied such transformations. Let $F \in C(\Omega), \Omega = I \times X \times U$, $C(\Omega)$ is the class of real-valued continuous functions and consider the mapping

$$\Lambda_p : F \in C(\Omega) \to \int_I F(t, x(t), u(t))dt \in R \tag{7.1}$$

where $p = (x(), u())$ be an admissible pair.

Problem 7.0.1. *Among the positive linear functionals on $C(\Omega)$ of the type Λ_p we seek the one for which the number $\Lambda_p(L)$ is a minimum, i.e.,*

$$\min I(p) = \min \Lambda_p(L) \tag{7.2}$$

$$\Lambda_p(\phi^f) = \Delta\phi, \quad \phi \in C'(B) \tag{7.3}$$

$$\Lambda_p(\psi_j) = 0, \quad j = 1, 2, \ldots, n, \psi \in \mathcal{D}(J^0) \tag{7.4}$$

$$\Lambda_p(g) = a_g, \quad g \in C_1(\Omega). \tag{7.5}$$

Similar transformations have been studied by Dantzig [30], Hernandez-Lerma and Lasserre [35], Pullan [59] in various applications. There are no efficient numerical techniques for such infinite dimensional LP problems. We wish to couple numerical analysis and symbolic Finsler geometry to try to obtain more efficient algorithms for feedback controls.

Multi-Agent Coordination via Connections

The algebra of connections can be viewed as an algebra of feedback controls. Various agents compute feedback controls in their local neighborhoods. When stitched together in an appropriate fashion, they yield possible feedback controls for distributed systems. We would like to study how to integrate the information from individual local sensors of system state, and what levels of communication are required among the agents to improve performance measures.

Bibliography

[1] R. Alur, D.L. Dill, *A theory of timed automata*, Theoretical Computer Science, 126, pp. 183-235, 1994.

[2] A. Agrachev, Y. Sachkov, *Control Theory from the Geometric Viewpoint* , Springer-Verlag, 2004.

[3] E.J. Anderson, P.Nash, *Linear Programming in Infinite-Dimensional Spaces*, Wiley, Chichester, UK, 1987.

[4] P.L. Antonelli, R.S. Ingarden, M. Matsumoto, *The Theory of Sprays and Finsler Spaces With Applications in Physics and Biology*, Kluwer Academic Publishers, 1994.

[5] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag; 2nd edition, 1989.

[6] U. Ascher, L Petzold, *Collocation software for boundary-value differential-algebraic equations*, SIAM J. Sci. Comput., 15, pp. 938-952, 1994.

[7] A.V. Balakrishnan, *On a new computing technique in optimal control*, SIAM Journal of Control., pp. 149-173, 1968.

[8] D. Bao, S. S. Chern, Z. Shen, *An Introduction to Riemann-Finsler Geometry*, Springer-Verlag, 2000.

[9] M. Bardi, I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Springer Verlag, 1997.

[10] R. Bellman,*Dynamic Programming*, Princeton Un iversity Press, Princeton, N.J., 1957.

[11] A. Bemporad, M Morari, *Control of systems integrating logic, dynamics, and constraints*, Automatica, 35(3), pp. 407-427, 1999.

[12] A. Bemporad, F. Borrelli, M. Morari, *Piecewise linear optimal controllers for hybrid systems*, In Proc. American Contr. Conf., Chicago, pp. 1190-1194, 2000.

[13] L. Berwald, *Uber Parallelubertragung in Raumen mit allgemeiner Massbestimmung*, Jahresber. Deutsch. Math.-Ver. 34, 213-220, 1926.

[14] J.T. Betts, *Survey of numerical methods for trajectory optimization*, J. Guidance, Control and Dynamics, Vol. 21, No 2, pp. 193-207, 1998.

[15] J.T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, 1st edition, 2001.

[16] A. Bloch, P. Crouch, J. Baillieul, J. Marsden, *Nonholonomic Mechanics and Control*, Springer Verlag, 2003.

[17] F. Borrelli, M. Baotic, A. Bemporad, M. Morari, *An efficient algorithm for computing the state feedback optimal control law for discrete time hybrid systems*, In Proc. American Contr. Conf., Denver, Colorado, 2003.

[18] M.S. Branicky, V.S. Borkar, and S.K. Mitter, *A unified framework for hybrid control: Model and optimal control theory*, IEEE Trans. Automatic Control, 43(1):31-45, January 1998.

[19] M. Broucke, S. Di Gennaro, M. Di Benedetto, and A. Sangiovanni-Vincentelli, *Efficient solution of optimal control problems using hybrid systems*, SIAM Journal on Control and Optimization, submitted December 2000.

[20] R. Bulirsch, H. J. Pesch, *The Maximum Principle, Bellman's Equation, and Caratheodary's work*, .

[21] C.G. Cassandras, D.L. Pepyne, Y. Wardi, *Optimal Control of a Class of Hybrid Systems*, IEEE Trans. on Automatic Control, AC-46, 3, pp. 398-415, 2001.

[22] C.G. Cassandras, K. Gokbayrak, *Optimal Control for Discrete Event and Hybrid Systems*, in Modeling, Control, and Optimization of Complex Systems, W. Gong and L. Shi, Editors, pp. 285-304, Kluwer, 2002.

[23] C. Caratheodary, *Calculus of Variations and Partial Differential Equations*, Translated by R.B. Dean and J.J. Brand, Holden day, 1965.

[24] C. Caratheodory, *Uber die diskontinuierlichen Losungen in der Variationsrechnung*, Dissertation, Gottingen, 1904, Collected Works, vol. 1, C. H. Beck, Munchen, pp. 3-79, 1954.

[25] C. Caratheodory, *Uber die starken Maxima and Minima bei einfachen Integralen*, Math. Ann. 62 (1906), 449-503. Collected Works, vol. 1, C. H. Beck, Munchen, pp. 80-142 1954.

[26] E. Cartan, *Les espaces de Finsler*, Actualites Sci. Indust. 79, 1934.

[27] V. Chandru, J. Hooker, *Optimization Methods for Logical Inference*, Wiley-Interscience; 1st edition, 1999.

[28] S. S. Chern, W. H. Chen, K. S. Lan, *Lectures on Differential Geometry*, World Scientific Pub Co Inc, 1999.

[29] S.S. Chern, *Finsler Geometry is just Riemannian Geometry without the Quadratic Restriction*, Notices Amer. Math. Soc., 43, pp. 959-963, 1996.

[30] G.B. Dantzig, *Linear Control Processes and Mathematical Programming*, J. SIAM Control, Vol 4, 56-60, 1966.

[31] P. Finsler, *Uber Kurven and Flachen in allgemeinen Raumen*, Dissertation, Gottingen, 1918. Reprinted by Birkhauser Verlag, Basel, 1951.

[32] W.H. Fleming, R.W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, 1975.

[33] N.I. Grachev and Yu.G. Evtushenko, *A Library of Programs for solving Optimal Control Problems*, U.S.S.R. Comput. Maths. Math. Phys., Vol.19, No.2, pp.99-119, 1980.

[34] T.A. Henzinger, *The theory of hybrid automata*, Proceedings of the 11th Annual Symposium on Logic in Computer Science (LICS), IEEE Computer Society Press, pp. 278-292. 1996.

[35] O. Hernandez-Lerma and J.B. Lasserre, *Approximation Schemes for Infinite Linear Programs*, SIAM J. Optimization, Vol 8, No. 4, pp. 973-988, 1998.

[36] M.R. Hestenes, *Calculus of Variations and Optimal Control Theory*, Wiley, New York, 1966.

[37] J. Imae, *A Riccati-equation based algorithm for continuous-time optimal control problems*, Optim. Control Appl. Meth., 19, pp. 299-313, 1998.

[38] R.G. Jeroslow, *Logic-Based Decision Support: Mixed Integer Model Formulation*, Annals of Discrete Mathematics, 40, North-Holland, 1989.

[39] V. Jurdjevic, *Geometric Control Theory*, Cambridge University Press, 1996.

[40] H. Keller, *Numerical methods for two-point boundary value problems*, Blaisdell, Waltham, MA, 1968.

[41] D. Kirk, *Optimal Control Theory: An Introduction*, Prentice-Hall, Englewood Cliffs, New Jersey, 1970.

[42] D. Kraft, *On converting optimal control problems into nonlinear programming problems*, Comp. Math. Prog., 15, pp. 261-280, 1985.

[43] W. Kohn, A. Nerode, J. Remmel, *Hybrid systems as Finsler manifolds: finite state control as approximation to connections*, Hybrid systems, II (Ithaca, NY, 1994), 294-321, Lecture Notes in Comput. Sci., 999, Springer, Berlin, 1995.

[44] W.Kohn, V. Brayman, and A.Nerode :*Control synthesis in Hybrid systems with Finsler dynamics*, Houston Journal of Mathematics, Vol.28, No. 2, 2002.

[45] W. Kohn, V. Brayman, P. Cholewinski and A. Nerode, *Control in Hybrid Systems*, International Journal of Hybrid Systems, Volume 3, No 2 & 3, 2003.

[46] M. Lentini, V. Pereyara, *An adaptive finite-difference solver for nonlinear two-point boundary problems with mild boundary layers*, SIAM J. Numer. Anal., 14, pp. 99-111, 1977.

[47] O. Hernandez-Lerma and J.B. Lasserre, *Approximation Schemes for Infinite Linear Programs*, SIAM J. Optimization, Vol 8, No. 4, pp. 973-988, 1998.

[48] D. Liberzon, *Switching in Systems and Control*, Birkhauser, Boston, MA, 2003.

[49] C. Lopez, E. Martinez, *Sub-Finslerian Metric Associated to an Optimal Control System*, SIAM J. Control Optim., Vol. 39, no. 3, pp.798-811, 2000.

[50] J. Lygeros, K.H. Johansson, S.N. Simic, J. Zhang, and S. Sastry, *Dynamical Properties of Hybrid Automata*, IEEE Transactions on Automatic Control, vol. 48, no. 1, pp. 2-18, 2003.

[51] M. Matsumoto, *Foundations of Finsler geometry and special Finsler spaces*, Kaiseicha Press, Otsu, 1986.

[52] R. Miron, D. Hrimiuc, H. Shimada, S. V. Sabau, *The Geometry of Hamilton and Lagrange Spaces*, Kluwer Academic Publishers, 2001.

[53] A. Nerode and W. Kohn, *Models for Hybrid Systems: Automata, Topologies, Controllability, Observability*, LNCS Hybrid Systems, 1993.

[54] H. Nijmeijer, A.J. van der Scahft, *Nonlinear Dynamical Control Systems*, Springer Verlag, New York, 1990.

[55] E. Polak, *Computational Methods in Optimization: A Unified Approach*, Academic Press, 1971.

[56] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer, New York, 1997.

[57] E. Polak, *A Historical Survey of Computational Methods in Optimal Control*, SIAM Review, Vol. 15, No.2, Part 2, pp. 553-584, 1973.

[58] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Michchenko, *The Mathematical Theory of Optimal Processes*, John-Wiley Interscience publishers 1962.

[59] M.C. Pullan, *An Algorithm for a Class of Continuous Linear Programs*, SIAM J. Control and Optimization, Vol. 31, no. 6, 1558-1577, 1993.

[60] R. Pytlak, *Numerical Methods for Optimal Control Problems with State Constraints*, Springer, 1st edition, 1999.

[61] B. Riemann, *Uber die Hypothesen, welche der Geometrie zugrunde liegen*, Habilitationsschrift, Gottingen, 1854. Collected Works, Dover Publications, NY, pp. 272-287, 1953.

[62] S. Roberts, J. Shipman, *Two-Point Boundary Value Problems: Shooting Methods*, American Elsevier, New York, 1972.

[63] J.E. Rubio, *Control and Optimization: The Linear Treatment of Nonlinear Problems*, Manchester University Press, Manchester, and John Wiley, New York and London, 1986.

[64] H. Rund, *The differential geometry of Finsler spaces*, Springer-Verlag, Berlin-Gottingen-Heidelberg, 1959.

[65] Z. Shen, *Differential Geometry of Spray and Finsler Spaces*, Kluwer Academic Publishers, 2001.

[66] G. Shilov, *Differential Geometry of Spray and Finsler Spaces*, Kluwer Academic Publishers, 2001.

[67] H. J. Sussmann, *An Introduction to the coordinate-free Maximum Principle*, in Geometry of Feedback and Optimal Control Theory, B. Jakubczyk, W. Respondek (eds.), pp. 463-557, Springer-Verlag, 1997.

[68] H. J. Sussmann, *Geometric Optimal Control*, in Mathematical Control Theory, J. Baillieul, Jan C. Willems (eds.), pp. 140-198, Springer-Verlag, 1998.

[69] J. Stoer, R. Bulirsch, R. Bartels, W. Gautschi, C. Witzgall, *Introduction to Numerical Analysis*, Springer-Verlag, 3 edition, 2002.

[70] K. Teo, C. Goh, K. Wong, *A unified Computational Approach to Optimal Control Problems*, Pitman Monographs and Surveys in Pure and Applied Mathematics, Wiley, New York, 1991.

[71] C. Udriste, *Finsler-Lagrange-Hamilton Structures Associated to Control Systems*, Proceedings of Conference on Finsler and Lagrange Geometries, Univ. Al.I.Cuza, Iasi, 26-31 August, 2001, edited by M. Anastasiei and P.L.Antonelli, pp.233-243 Kluwer Academic Publishers, 2003.

[72] O. Von Stryk, *Numerical solution of optimal control problems by direct collocation*, in Optimal Control and Variational Calculus, R. Bulirsch, A. Miele, J. Stoer, and K. Well, eds., Birkhauser, Basel, Germany, 1993.

[73] J. Warga, *Optimal Control of Differential and Functional equations*, Academic Press, New York, 1972.

[74] H.P. Williams, *Model Building in Mathematical Programming*, John Wiley & Sons; 4th edition, 1999.

[75] V. Yen, M. Nagurka, *Optimal control of linearly constrained systems via state parametrization*, Opt. Cont. Appl. and Meth., 13, pp. 155-167, 1992.

[76] L.C. Young, *Lectures on Calculus of Variations and Optimal Control Theory*, Philadelphia: Saunders, 1969.