THE CALCULUS OF VARIATIONS
LAGRANGIAN MECHANICS AND OPTIMAL CONTROL

AARON NELSON

ABSTRACT. The calculus of variations addresses the need to optimize certain quantities over sets of functions. We begin by defining the notion of a functional and its role in performing calculus on variations of curves. In order to identify those functions which are extremals of a functional, we establish the first variation and the Euler-Lagrange Equations. A formulation of the second variation follows, providing a means of identifying different types of extremals. Both the first and second variations are illustrated using several examples. We then introduce the calculus of variations as it applies to classical mechanics, resulting in the Principle of Stationary Action, from which we develop the foundations of Lagrangian mechanics. Finally, we examine an extension of the calculus of variations in optimal control. We conclude with Pontryagin's Maximum Principle and its applications in control theory.

CONTENTS

1. INTRODUCTION

On the first of January, 1697, Swiss mathematician Johann Bernoulli published a challenge problem as his New Year's gift to the mathematical world. The problem, which he dubbed the "brachistochrone" problem from the Greek words *brachistos* and *chronos*, meaning "shortest" and "time," respectively, stated:

> *To determine the curved line joining two given points, situated at different distances from the horizontal and not in the same vertical line, along which a mobile body, running down by its own weight and starting to move from the upper point, will descend most quickly to the lowest point.* [Ric96]

While not entirely new, having been attempted by Galileo in 1638, Bernoulli himself was the first to provide a correct solution. Several solutions were submitted, including those of famed figures

Undergraduate honors thesis, May 2012. Faculty advisor: Dr. Maria Gordina, University of Connecticut Department of Mathematics, Storrs, CT 06269.

such as Leibniz, Newton, and L'Hospital. However, it was Johann's older brother, Jakob, whose solution eventually laid the groundwork for what would become known as the calculus of variations. Jakob noticed that the problem was of a new type: the variables were functions. Indeed, the brachistochrone problem is an example of an instance in which one wishes to optimize a function whose domain is a space of functions. Such a function is called a functional, the focal point of the calculus of variations.

More generally, a functional is defined as "any mapping from the space of curves to the real numbers" [Arn78]. Before we delve into defining functionals and the necessary conditions for optimization, however, we will establish some basic notation:

(1) Let $f : \mathbb{R} \to \mathbb{R}^k$. Then, we denote by $\dot{f}$ the time derivative, $\mathrm{d}f/\mathrm{d}t$, or its equivalent in the context of a given problem. Similarly, we use $\ddot{f}$ to denote $\mathrm{d}^2 f/\mathrm{d}t^2$.
(2) For $q_i : \mathbb{R} \to \mathbb{R}$ we use $\mathbf{q} : \mathbb{R} \to \mathbb{R}^k$ to denote $\mathbf{q}(t) = (q_1(t), q_2(t), ..., q_k(t))$. Likewise, $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ denote $\dot{\mathbf{q}}(t) = (\dot{q}_1(t), \dot{q}_2(t), ..., \dot{q}_k(t))$ and $\ddot{\mathbf{q}}(t) = (\ddot{q}_1(t), \ddot{q}_2(t), ..., \ddot{q}_k(t))$, respectively. In the case of $k = 1$, we simply write q, $\dot{q}$, and $\ddot{q}$.
(3) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, we write $\mathbf{xy}$ to denote the product $\mathbf{x}^T\mathbf{y} = (x_1 y_1, x_2 y_2, ..., x_k y_k) \in \mathbb{R}^k$.
(4) We say a function $f : U \to \mathbb{R}^k$ is of class C^n if it is at least n-times continuously differentiable on the open set $U \subset \mathbb{R}$.

With this notation we can now begin to formally define a functional.

Definition 1.1. Let $\mathbf{u} \in C^1$ such that $\mathbf{u} : (t_1, t_2) \to U \subset \mathbb{R}$ and $\mathrm{L} : U \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ be of class C^1. Then, the quantity $\mathrm{J} : C^1 \to \mathbb{R}$ given by

$$\mathrm{J}[\mathbf{u}] = \int_{t_1}^{t_2} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)\ \mathrm{d}t$$

is called an *objective functional*, which we refer to simply as a functional.

A functional has as its domain a set of functions and as its range the real numbers. The function $\mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}, t)$ in the integrand is known as the *Lagrangian*. Often, quantities related to dynamical systems can be expressed using a Lagrangian. Thus, it is reasonable to suppose that if we construct the appropriate Lagrangian for a specific system, the range of the functional will be a quantity which we may want to optimize over a given set of functions. This is the basis for the study of the calculus of variations. In this paper we will explore the fundamentals of the calculus of variations. We begin by establishing the first variation and the Euler-Lagrange equations before introducing the second variation through an analysis of the quadratic functional. Both the first and second variations are illustrated through several well known examples, including geodesics, the minimal surface of revolution, and the bractistochrone. We then utilize the first variation in the formulation of Lagrangian mechanics via Hamilton's Principle of Stationary Action. We conclude with a discussion of basic optimal control theory and Pontryagin's Maximum Principle as an extension of the calculus of variations.

2. The First Variation

Recall from calculus that, given a function $f : \mathbb{R}^k \to \mathbb{R}$, a process of optimization was determining the points at which local maxima and minima of f were obtained. This involved two steps which utilized derivatives of the function. First, candidates for local maxima and minima were found by locating critical points, values for which the gradient, $\nabla f = (\partial f/\partial x_1, \partial f/\partial x_2, ..., \partial f/\partial x_k,)$, is zero (the first derivative test). Critical points were then classified as local maxima, minima, or saddle points based on a concavity argument established using the second partial derivatives of f (the second derivative test). In order to optimize a functional we wish to develop functional analogues to these first and second derivative tests. For now, we will focus on finding the critical functions which will be our candidates for local maxima and minima. Our treatment of the subject is inspired by the work of [Olv12].

We turn first to the task of finding the functional gradient, denoted $\nabla \mathrm{J}[\mathbf{u}]$, which will be the centerpiece of our first derivative test for functionals. We treat the functional gradient as a directional derivative,

$$\langle \nabla \mathrm{J}[\mathbf{u}], \mathbf{v} \rangle = \frac{\mathrm{d}}{\mathrm{d}\lambda} \mathrm{J}[\mathbf{u} + \lambda \mathbf{v}] \Big|_{\lambda=0} \ ,$$

where $\lambda \in \mathbb{R}$. The function $\mathbf{v}$ representing the direction of the derivative is called a *variation* of the function $\mathbf{u}$ and the resulting functional gradient is the *first variation* of $\mathrm{J}[\mathbf{u}]$. The inner product is the standard L^2 inner product for real functions, $\langle f, g \rangle = \int f(x)\, g(x)\, \mathrm{d}x$. We are now in a position to explicitly define the functional gradient.

Proposition 2.1 (The First Variation)**.** *Let* $\mathbf{u} : [t_1, t_2] \to \mathbb{R}^k$ *be a function of class* C^2 *satisfying the boundary conditions* $\mathbf{u}(t_1) = \mathbf{x_1}$*,* $\mathbf{u}(t_2) = \mathbf{x_2}$*. Then, the first variation of the functional* $\mathrm{J}[\mathbf{u}]$ *is given by the functional gradient,*

$$\nabla \mathrm{J}[\mathbf{u}] = \frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \right) \ , \tag{1}$$

where $\mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \in C^2$ *is the associated Lagrangian.*

Proof. Let $\mathbf{v} \in C^2$ be a variation of $\mathbf{u}$. Then, $\mathbf{v} : [t_1, t_2] \to \mathbb{R}^k$ satisfies the boundary conditions $\mathbf{u}(t_1) + \lambda \mathbf{v}(t_1) = \mathbf{x_1}$ and $\mathbf{u}(t_2) + \lambda \mathbf{v}(t_2) = \mathbf{x_2}$. By definition, the first variation of $\mathrm{J}[\mathbf{u}]$ is given by the inner product

$$\begin{aligned}
\langle \nabla \mathrm{J}[\mathbf{u}], \mathbf{v} \rangle &= \frac{\mathrm{d}}{\mathrm{d}\lambda} \mathrm{J}[\mathbf{u} + \lambda \mathbf{v}] \Big|_{\lambda=0} = \frac{\mathrm{d}}{\mathrm{d}\lambda} \left[\int_{t_1}^{t_2} \mathrm{L}(\mathbf{u} + \lambda \mathbf{v}, \dot{\mathbf{u}} + \lambda \dot{\mathbf{v}}, t)\ \mathrm{d}t \right] \Big|_{\lambda=0} \\
&= \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}\lambda} \mathrm{L}(\mathbf{u} + \lambda \mathbf{v}, \dot{\mathbf{u}} + \lambda \dot{\mathbf{v}}, t) \Big|_{\lambda=0} \mathrm{d}t
\end{aligned}$$

for $\lambda \in \mathbb{R}$. Applying the chain rule and evaluating at $\lambda = 0$ yields

$$\begin{aligned}
\langle \nabla \mathrm{J}[\mathbf{u}], \mathbf{v} \rangle &= \int_{t_1}^{t_2} \left[\mathbf{v} \frac{\partial [\mathrm{L}(\mathbf{u} + \lambda \mathbf{v}, \dot{\mathbf{u}} + \lambda \dot{\mathbf{v}}, t)]}{\partial (\mathbf{u} + \lambda \mathbf{v})} + \dot{\mathbf{v}} \frac{\partial [\mathrm{L}(\mathbf{u} + \lambda \mathbf{v}, \dot{\mathbf{u}} + \lambda \dot{\mathbf{v}}, t)]}{\partial (\dot{\mathbf{u}} + \lambda \dot{\mathbf{v}})} \right] \Big|_{\lambda=0} \mathrm{d}t \\
&= \int_{t_1}^{t_2} \left[\mathbf{v} \frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) + \dot{\mathbf{v}} \frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \right] \mathrm{d}t \ .
\end{aligned}$$

We eliminate the $\dot{\mathbf{v}}$ term in the integrand by integrating the second summand by parts:

$$\int_{t_1}^{t_2} \dot{\mathbf{v}} \frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)\, \mathrm{d}t = (\mathbf{v})\, \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \mathbf{v} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \right) \mathrm{d}t \ . \tag{2}$$

Since $\mathbf{u} + \lambda \mathbf{v}$ and $\mathbf{u}$ satisfy the same boundary conditions, we see that $\mathbf{v}(t_1) = \mathbf{v}(t_2) = 0$. Thus, substituting equation (2) back into the inner product gives us

$$\langle \nabla \mathrm{J}[\mathbf{u}], \mathbf{v} \rangle = \int_{t_1}^{t_2} \left[\mathbf{v} \frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) - \mathbf{v} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \right) \right] \mathrm{d}t$$

$$= \int_{t_1}^{t_2} \mathbf{v} \left[\frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \right] \mathrm{d}t \ .$$

Equating this result with the integral form of the inner product yields

$$\int_{t_1}^{t_2} \mathbf{v} \, \nabla \mathrm{J}[\mathbf{u}] \, \mathrm{d}t = \int_{t_1}^{t_2} \mathbf{v} \left[\frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \right] \mathrm{d}t \ .$$

Because $\mathbf{v}$ is a random variation of $\mathbf{u}$, this equality must hold for all choices of $\mathbf{v}$. Thus, we obtain

$$\nabla \mathrm{J}[\mathbf{u}] = \frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \ ,$$

as desired. □

Much like the first derivative test in calculus, a function in the domain of a functional is a critical function if it causes the functional gradient to vanish. Thus, our set of critical functions are stable points of the first variation. This leads us to our first theorem.

Theorem 2.2 (The Necessary Condition of Euler-Lagrange)**.** *Suppose* $\mathbf{u}$ *and* $\mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right)$ *satisfy the hypotheses of Proposition 2.1. Then,* $\mathbf{u}$ *is a critical function of* $\mathrm{J}[\mathbf{u}]$ *if and only if*

$$\frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \ . \tag{3}$$

Proof. In order to be a critical function, $\mathbf{u}$ must satisfy $\nabla \mathrm{J}[\mathbf{u}] = 0$. From Proposition 2.1, we have that

$$\nabla \mathrm{J}[\mathbf{u}] = \frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \ .$$

Equating this to zero immediately yields the desired result,

$$\frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \ .$$

□

The set of equations (3) are refered to as the *Euler-Lagrange Equations*. Since the majority of problems in the calculus of variations involve finding local minima of functionals, satisfying these equations is a condition we will require of solutions. Indeed, the Euler-Lagrange Equations often yield only the desired solutions to these problems, without the need for further analysis. Since we will use them frequently, we abbreviate the Euler-Lagrange Equations by

$$\frac{\partial \mathrm{L}}{\partial \mathbf{u}} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{u}}} \right)$$

for the remainder of the text.

Example 2.3 (Planar Geodesic)**.** Our first example of a variational problem is the planar geodesic: given two points lying in a plane, we want to determine the path of shortest distance between them. Let (x_1, y_1), (x_2, y_2) denote two points in the plane and $u : [x_1, x_2] \to \mathbb{R}$ be of class C^2 satisfying the boundary conditions $u\left(x_1\right) = y_1$, $u\left(x_2\right) = y_2$. Our goal is to determine which function u has minimal arclength between these points. Naturally, we expect this function to be linear.

First, we must formulate the arclength Lagrangian. From calculus, we know that the arclength of a curve, which we denote $|u|$, is given by the sum

$$|u| = \lim_{n\to\infty} \sum_{i=1}^{n} |P_i - P_{i-1}| \quad , \tag{4}$$

where each $P_i = (x_i, y_i)$ is a point on the curve between the boundaries. For $|P_i - P_{i-1}|$, the arclength is calculated using the linear distance formula:

$$|P_i - P_{i-1}| = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2} \quad .$$

By the Mean Value Theorem, we also have that $\Delta y_i = \dot{u}\,(x_i^*)\,\Delta x_i$ for some $x_i^* \in (x_{i-1}, x_i)$. Thus,

$$|P_i - P_{i-1}| = \sqrt{(\Delta x_i)^2 \left(1 + \dot{u}\,(x_i^*)^2\right)} = \Delta x_i \sqrt{1 + \dot{u}\,(x_i^*)^2} \quad .$$

Substiting this into equation (4) gives us

$$|u| = \lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i \sqrt{1 + \dot{u}\,(x_i^*)^2} = \int_{x_1}^{x_2} \sqrt{1 + \dot{u}\,(x)^2}\,\mathrm{d}x \quad .$$

Notice that this result is a functional and the arclength Lagrangian is the integrand. Thus, to determine the planar geodesic we want to minimize the functional

$$\mathrm{J}[u] = \int_{x_1}^{x_2} \mathrm{L}\,(u, \dot{u}, x)\ \mathrm{d}x$$

with Lagrangian given by

$$\mathrm{L}(u, \dot{u}, x) = \sqrt{1 + \dot{u}^2} \quad .$$

From Theorem 2.2, we require that any minimizer of $\mathrm{J}[u]$ satisfy the Euler-Lagrange Equations. In this case, there is only one:

$$\frac{\partial \mathrm{L}}{\partial u} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial \mathrm{L}}{\partial \dot{u}}\right) \quad .$$

Taking each side of the equation separately, we have

$$\frac{\partial \mathrm{L}}{\partial u} = 0 \quad ,$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial \mathrm{L}}{\partial \dot{u}}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\dot{u}}{\sqrt{1 + \dot{u}^2}}\right) = \frac{\ddot{u}}{(1 + \dot{u}^2)^{3/2}} \quad .$$

Thus, any minimizing curve satisfies the differential equation

$$\frac{\ddot{u}}{(1 + \dot{u}^2)^{3/2}} = 0 \quad . \tag{5}$$

Clearly, equation (5) is equivalent to $\ddot{u} = 0$. The function which satisfies this equation is necessarily linear. Explicitly, the shortest path between two points (x_1, y_1), (x_2, y_2) in the plane is the straight line given by

$$u\,(x) = \left(\frac{y_2 - y_1}{x_2 - x_1}\right) x + \frac{y_1 x_2 - x_1 y_2}{x_2 - x_1} \quad . \tag{6}$$

This confirms our initial intuition, although further work is necessary to complete the proof that this critical function is actually a minima and not a maxima or saddle point. Later, we will develop the tools necessary to do this.

Let us look now at an alternative method for determining the critical functions of $\mathrm{J}[\mathbf{u}]$. For this, we introduce Poisson's variables and Hamilton's Equations.

Definition 2.4. Let $\mathbf{u} \in C^2$ and $\mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)$ be a Lagrangian. We define $\mathbf{p} \in C^2$ to be *Poisson's variables*,

$$\mathbf{p} = \frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \quad , \tag{7}$$

and the quantity

$$\mathrm{H}(\mathbf{u}, \mathbf{p}, t) = \mathbf{p}\dot{\mathbf{u}} - \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)$$

to be the *Hamiltonian*, which is the Legendre transform of the Lagrangian. Then, *Hamilton's Equations* are given by

$$\dot{\mathbf{u}} = \frac{\partial}{\partial \mathbf{p}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) \quad ,$$
$$\dot{\mathbf{p}} = -\frac{\partial}{\partial \mathbf{u}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) \quad .$$

Further information on the Legendre transform of the Lagrangian is available in [Arn78]. Having defined the Hamiltonian, we are now in a position to introduce our second theorem.

Theorem 2.5. *Let* $\mathbf{u}, \mathbf{p} \in C^2$ *satisfy equations (7),* $\mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)$ *be a Lagrangian, and* $\mathrm{H}(\mathbf{u}, \mathbf{p}, t)$ *be the associated Hamiltonian. Then, Hamilton's Equations and the Euler-Lagrange Equations are equivalent.*

Proof. The total derivative of the Hamiltonian is given by the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = \dot{\mathbf{u}} \frac{\partial}{\partial \mathbf{u}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) + \dot{\mathbf{p}} \frac{\partial}{\partial \mathbf{p}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) + \frac{\partial}{\partial t} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) \quad .$$

Using the definition of the Hamiltonian, we also have that this total derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathbf{p}\dot{\mathbf{u}} - \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \right) = -\dot{\mathbf{u}} \frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) + \dot{\mathbf{p}}\dot{\mathbf{u}} - \frac{\partial}{\partial t} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \quad .$$

Since these two expressions must be equal, we necessarily have

$$\frac{\partial}{\partial \mathbf{u}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = -\frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \quad ,$$
$$\frac{\partial}{\partial \mathbf{p}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = \dot{\mathbf{u}} \quad ,$$
$$\frac{\partial}{\partial t} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = -\frac{\partial}{\partial t} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t) \quad .$$

Clearly, the second of these equations is true since it is one of Hamilton's Equations. Thus, we need only show that

$$\frac{\partial}{\partial \mathbf{u}} \mathrm{H}(\mathbf{u}, \mathbf{p}, t) = -\frac{\partial}{\partial \mathbf{u}} \mathrm{L}(\mathbf{u}, \dot{\mathbf{u}}, t)$$

if and only if $\mathbf{u}$ satisfies both Hamilton's Equations and the Euler-Lagrange Equations. First, assume $\mathbf{u}$ satisfies Hamilton's Equations. Then, we see by substitution that

$$\dot{\mathbf{p}} = \frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \quad .$$

Second, assume $\mathbf{u}$ satisfies the Euler-Lagrange Equations. Then, taking the time derivative of equations (7),

$$\dot{\mathbf{p}} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial}{\partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \right) \quad ,$$

and applying the Euler-Lagrange Equations gives us

$$\dot{\mathbf{p}} = \frac{\partial}{\partial \mathbf{u}} \mathrm{L}\left(\mathbf{u}, \dot{\mathbf{u}}, t\right) \quad .$$

Since both cases yield the same result, we conclude that Hamilton's Equations and the Euler-Lagrange Equations are equivalent, as desired. □

Due to the symmetry of Hamilton's Equations, they will lead us to some convenient ways of determining critical functions of $\mathrm{J}[\mathbf{u}]$. Since we will use them frequently, we abbreviate the Lagrangian and Hamiltonian by L and H, respectively, and Hamilton's Equations by

$$\dot{\mathbf{u}} = \frac{\partial \mathrm{H}}{\partial \mathbf{p}}$$
$$\dot{\mathbf{p}} = -\frac{\partial \mathrm{H}}{\partial \mathbf{u}}$$

for the remainder of the text.

Example 2.6 (The Brachistochrone)**.** Let us now consider the example which we introduced as the origin of the calculus of variations, the brachistochrone problem. A bead, initially at rest, is allowed to slide down a wire with endpoints fixed at different heights, not directly above one another, due solely to its own weight. The length and shape of the wire are not specified. We want to determine the shape of the wire which minimizes the time it takes for the bead to travel its length.

Without loss of generality, let $(0, 0)$, (x_1, y_1) be the endpoints of a wire such that $x_1 > 0$ and $y_1 > 0$ (we define our coordinate system so that the downward direction is $+y$). Moreover, let $u : [0, x_1] \to \mathbb{R}$ be of class C^2 satisfying the boundary conditions $u\left(0\right) = 0$, $u\left(x_1\right) = y_1$. Our goal is to determine which function u describes the path of the bead as it slides along the wire in the least amount of time. Unlike in our first example, the result is far from intuitive.

First, we must formulate the time Lagrangian. From Example 2.3, we know that the arclength is given by

$$|u| = \lim_{n \to \infty} \sum_{i=1}^{n} \Delta x_i \sqrt{1 + \dot{u}\left(x_i^*\right)^2} \quad ,$$

from which we obtain the arclength element, ΔS:

$$\Delta S = \Delta x \sqrt{1 + \dot{u}^2} \quad .$$

Since the instantaneous speed of the bead along the curve is given by

$$v\left(x\right) = \frac{\Delta S}{\Delta t} \; ,$$

rearranging and substitution yields

$$\Delta t = \frac{\Delta S}{v} = \Delta x \frac{\sqrt{1+\dot{u}^2}}{v} \ . \tag{8}$$

To find v, we use the principle of energy conservation. Notice that $u(0) = 0$ and, since the bead starts at rest, $v(0) = 0$. This means that at $x = 0$ the kinetic energy,

$$T = \frac{1}{2}\mathrm{m}v^2 \ ,$$

and potential energy,

$$U = -\mathrm{mg}u \ ,$$

of the system are both zero. Thus, the total energy, $T + U$, is also zero, and by conservation of energy we have that

$$T + U = 0 \quad \Rightarrow \quad \frac{1}{2}\mathrm{m}v^2 = \mathrm{mg}u \quad \Rightarrow \quad v = \sqrt{2\mathrm{g}u} \ .$$

Substituting this quantity into equation (8) gives us that the time Lagrangian is

$$\mathrm{L}(u, \dot{u}, x) = \frac{\Delta t}{\Delta x} = \sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}} \ .$$

Thus, to find the curve which minimizes the time it takes for the bead to travel from $(0,0)$ to (x_1, y_1), we want to minimize the functional

$$\mathrm{J}[u] = \int_0^{x_1} \sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}}\,\mathrm{d}x \ .$$

From Theorem 2.2, we require that any minimizer of $\mathrm{J}[u]$ satisfy the Euler-Lagrange Equations. Once again, there is only one:

$$\frac{\partial \mathrm{L}}{\partial u} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial \mathrm{L}}{\partial \dot{u}}\right) \ .$$

Thus, we have

$$\frac{\partial \mathrm{L}}{\partial u} = \frac{\partial}{\partial u}\left(\sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}}\right) = \frac{-\mathrm{g}\sqrt{1+\dot{u}^2}}{(2\mathrm{g}u)^{3/2}} \ ,$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial \mathrm{L}}{\partial \dot{u}}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial}{\partial \dot{u}}\left(\sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}}\right)\right] = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\dot{u}}{\sqrt{2\mathrm{g}u\,(1+\dot{u}^2)}}\right) = \frac{2\mathrm{g}u\ddot{u} - \mathrm{g}\dot{u}^2\,(1+\dot{u}^2)}{(2\mathrm{g}u\,(1+\dot{u}^2))^{3/2}} \ .$$

Any minimizing curve must therefore satisfy the differential equation

$$\frac{2\mathrm{g}u\ddot{u} - \mathrm{g}\dot{u}^2\,(1+\dot{u}^2)}{(2\mathrm{g}u\,(1+\dot{u}^2))^{3/2}} = \frac{-\mathrm{g}\sqrt{1+\dot{u}^2}}{(2\mathrm{g}u)^{3/2}} \ ,$$

which can be simplified to

$$2u\ddot{u} + \dot{u}^2 + 1 = 0 \ . \tag{9}$$

This second order differential equation is rather complicated and difficult to solve. Luckily, Hamilton's Equations yield a much simpler solution. Using Hamilton's Equations, we can write the Hamiltonian as

$$\mathrm{H} = p\dot{u} - \mathrm{L} = \dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} \ \ .$$

Since L does not depend explicitly on x, we have $\partial \mathrm{L}/\partial x = 0$. This necessarily means that

$$\frac{\mathrm{dH}}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L}\right) = 0 \quad \Rightarrow \quad \dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} = C \ \ ,$$

for some constant C. Thus,

$$\begin{aligned}
&\dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} = C \\
&\Rightarrow \quad \dot{u}\left(\frac{\dot{u}}{\sqrt{2\mathrm{g}u\left(1+\dot{u}^2\right)}}\right) - \sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}} = C \\
&\Rightarrow \quad \frac{-1}{\sqrt{2\mathrm{g}u\left(1+\dot{u}^2\right)}} = C \ \ .
\end{aligned}$$

From this we conclude that any minimizing curve must also satisfy the differential equation

$$\left(1+\dot{u}^2\right)u = k^2 \ , \qquad \text{where } \ k^2 = \frac{-1}{2\mathrm{g}C^2} \ \ . \tag{10}$$

This equation is much more familiar than equation (9). Rearranging terms, we can express equation (10) as

$$\dot{u}^2 = \frac{k^2 - u}{u} \ \ ,$$

which resembles the more general differential equation

$$\dot{y}^2 = \frac{2R - y}{y} \ \ . \tag{11}$$

Equation (11) is separable and can be solved by direct integration using the trigonometric substitution $y = R\left(1 - \cos t\right)$:

$$\begin{aligned}
&\dot{y} = \frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{\frac{2R-y}{y}} \\
&\Rightarrow \quad \int \sqrt{\frac{y}{2R-y}}\ \mathrm{d}y = \int \mathrm{d}x \\
&\Rightarrow \quad \int \sqrt{\frac{(1-\cos t)}{(1+\cos t)}}\ R\sin(t)\ \ \mathrm{d}t = x + C \\
&\Rightarrow \quad \int R\left(1-\cos t\right)\ \mathrm{d}t = x + C \\
&\Rightarrow \quad x = R\left(t - \sin t\right) - C
\end{aligned}$$

where C is a constant of integration. This solution, parametrized by the equations

$$\begin{cases} x = R\,(t - \sin t) \\ y = R\,(1 - \cos t) \end{cases} \quad ,$$

where C is taken to be zero for simplicity, is the cycloid formed by revolutions of a circle of radius R. Thus, if we let $R \equiv \frac{1}{2}k^2$, the solution to equation (10) is the cycloid given by

$$\begin{cases} x = \frac{1}{2}k^2\,(t - \sin t) \\ u = \frac{1}{2}k^2\,(1 - \cos t) \end{cases} \quad . \tag{12}$$

Explicitly, the path between the points $(0, 0)$, (x_1, y_1) which minimizes the bead's time of travel is given by the parametrization in equation (12) with constant k and time t_1 determined from the boundary condition $u\,(x_1) = y_1$:

$$x_1 = \frac{1}{2}k^2\,(t_1 - \sin t_1) \quad ,$$
$$y_1 = \frac{1}{2}k^2\,(1 - \cos t_1) \quad .$$

The brachistochrone curve is therefore a cycloid. This result is not immediately obvious, illustrating the utility of the calculus of variations.

Notice that the derivation of equation (10) required that the Hamiltonian be independent of the variable x. In general, if $\mathrm{dH}/\mathrm{d}t = 0$ we call the Hamiltonian a *conserved* quantity. In such cases, the Hamiltonian is the integral solution to a differential equation. For this reason, we refer to it as a *first integral* solution. Problems in which the Hamiltonian is a first integral solution can be somewhat easier to solve using Hamilton's Equations than the Euler-Lagrange Equations. Because of this, we would like to have a way of determining when the Hamiltonian is a conserved quantity.

Proposition 2.7. *Let* $\mathbf{u}, \mathbf{p}, \mathrm{L},$ *and* H *satisfy the hypotheses of Theorem 2.5. Then, the Hamiltonian is a first integral solution satisfying* $\mathrm{dH}/\mathrm{d}t = 0$ *if and only if* $\partial\mathrm{L}/\partial t = 0$.

Proof. Recall from the proof of Theorem 2.5 that the total derivative of the Hamiltonian is given by

$$\frac{\mathrm{dH}}{\mathrm{d}t} = -\dot{\mathbf{u}}\frac{\partial \mathrm{L}}{\partial \mathbf{u}} + \dot{\mathbf{p}}\dot{\mathbf{u}} - \frac{\partial \mathrm{L}}{\partial t}$$

and that $\dot{\mathbf{p}} = \partial\mathrm{L}/\partial\mathbf{u}$. Combining these results yields

$$\frac{\mathrm{dH}}{\mathrm{d}t} = -\dot{\mathbf{p}}\dot{\mathbf{u}} + \dot{\mathbf{p}}\dot{\mathbf{u}} - \frac{\partial \mathrm{L}}{\partial t} = -\frac{\partial \mathrm{L}}{\partial t} \quad .$$

For the Hamiltonian to be a first integral solution, we require that $\mathrm{dH}/\mathrm{d}t = 0$. Clearly, this only holds if $\partial\mathrm{L}/\partial t = 0$, as desired. □

Although both the Euler-Lagrange Equations and Hamilton's Equations will yield the same critical functions of $\mathrm{J}[\mathbf{u}]$, in practice one is usually simpler and easier to work with than the other. We will generally use the simpler of the two and omit the other whenever possible.

3. The Second Variation

As we have already discussed, the first variation of a functional allows us to determine its critical functions, those which are candidates for optimization. Often, it only gives us the minimizers we are looking for, but we cannot fully prove that a particular critical function is actually a minimizer with the first variation alone. To do that, we will need to develop more tools. To this end, we now shift our attention toward determining the functional analogue to the second derivative test in calculus. Let us begin by looking at the first few terms in the Taylor expansion of $\mathrm{J}[\mathbf{u}+\lambda\mathbf{v}]$, where $\mathbf{v}$ is a variation of $\mathbf{u}$:

$$\mathrm{J}[\mathbf{u}+\lambda\mathbf{v}] = \mathrm{J}[\mathbf{u}] + \lambda\,\langle\nabla\mathrm{J}[\mathbf{u}],\mathbf{v}\rangle + \frac{1}{2}\lambda^2\,\mathrm{Q}[\mathbf{u},\mathbf{v}] + \ldots\ .$$

If $\mathbf{u}$ is a critical function, we know that $\langle\nabla\mathrm{J}[\mathbf{u}],\mathbf{v}\rangle$ vanishes by Theorem 2.2, so the behavior of the functional is relient upon the second derivative terms, $\mathrm{Q}[\mathbf{u},\mathbf{v}]$. These terms are called the *second variation* of $\mathrm{J}[\mathbf{u}]$ and are given by the second functional derivative,

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\,\mathrm{J}[\mathbf{u}+\lambda\mathbf{v}]\,\Big|_{\lambda=0}\ .$$

Proposition 3.1 (The Second Variation)**.** *Let* $\mathbf{u}$, $\mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right)$ *satisfy the hypotheses of Proposition 2.1 and* $\mathbf{v}\in C^2$ *be a variation of* $\mathbf{u}$*. Then, the second variation of the functional* $\mathrm{J}[\mathbf{u}]$ *is given by*

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \int_{t_1}^{t_2}\left[\mathbf{v}^2\,\frac{\partial^2}{\partial\mathbf{u}^2}\,\mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + 2\mathbf{v}\dot{\mathbf{v}}\,\frac{\partial^2}{\partial\mathbf{u}\partial\dot{\mathbf{u}}}\mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + \dot{\mathbf{v}}^2\,\frac{\partial^2}{\partial\dot{\mathbf{u}}^2}\,\mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right)\right]\,\mathrm{d}t\ . \quad (13)$$

Proof. By definition, the second variation of $\mathrm{J}[\mathbf{u}]$ is given by the derivative

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\,\mathrm{J}[\mathbf{u}+\lambda\mathbf{v}]\,\Big|_{\lambda=0} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\langle\nabla\mathrm{J}[\mathbf{u}],\mathbf{v}\rangle\,\Big|_{\lambda=0}$$

for $\lambda\in\mathbb{R}$. From the proof of Proposition 2.1, we know that

$$\langle\nabla\mathrm{J}[\mathbf{u}],\mathbf{v}\rangle = \int_{t_1}^{t_2}\left[\mathbf{v}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)} + \dot{\mathbf{v}}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)}\right]\Big|_{\lambda=0}\,\mathrm{d}t\ ,$$

so we have

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \int_{t_1}^{t_2}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\mathbf{v}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)} + \dot{\mathbf{v}}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)}\right]\Big|_{\lambda=0}\,\mathrm{d}t\ . \quad (14)$$

Applying the chain rule to each summand yields

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\mathbf{v}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)}\right)\\
&\quad= \mathbf{v}^2\,\frac{\partial^2\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)^2} + \mathbf{v}\dot{\mathbf{v}}\,\frac{\partial^2\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)}\ ,\\
&\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\dot{\mathbf{v}}\,\frac{\partial\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)}\right)\\
&\quad= \mathbf{v}\dot{\mathbf{v}}\,\frac{\partial^2\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\mathbf{u}+\lambda\mathbf{v}\right)\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)} + \dot{\mathbf{v}}^2\,\frac{\partial^2\left[\mathrm{L}\left(\mathbf{u}+\lambda\mathbf{v},\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}},t\right)\right]}{\partial\left(\dot{\mathbf{u}}+\lambda\dot{\mathbf{v}}\right)^2}\ ,
\end{aligned}$$

so substituting this back into equations (14) and evaluating at $\lambda = 0$ gives us

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \int_{t_1}^{t_2} \left[\mathbf{v}^2 \, \frac{\partial^2}{\partial \mathbf{u}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + \dot{\mathbf{v}}^2 \, \frac{\partial^2}{\partial \dot{\mathbf{u}}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) \right] \, \mathrm{d}t$$

as desired. □

Much like for a critical point of a function in calculus, a critical function $\mathbf{u}$ is a minimizer of the functional $\mathrm{J}[\mathbf{u}]$ if it makes the second variation positive definite for any nonzero variation $\mathbf{v}$. This is the basis of our next theorem.

Theorem 3.2. *Suppose* $\mathbf{u}$*,* $\mathbf{v}$*, and* $\mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right)$ *satisfy the hypotheses of Proposition 3.1 such that* $\mathbf{u}$ *is a critical function of* $\mathrm{J}[\mathbf{u}]$*. Then,* $\mathbf{u}$ *is a local minimum of* $\mathrm{J}[\mathbf{u}]$ *if and only if*

$$\mathbf{v}^2 \, \frac{\partial^2}{\partial \mathbf{u}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + \dot{\mathbf{v}}^2 \, \frac{\partial^2}{\partial \dot{\mathbf{u}}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) > 0 \tag{15}$$

for every nonzero variation $\mathbf{v}$*.*

Proof. In order to be a minimizer, $\mathbf{u}$ must satisfy $\mathrm{Q}[\mathbf{u},\mathbf{v}] > 0$ for any nonzero variation $\mathbf{v}$. From Proposition 3.1, we have that

$$\mathrm{Q}[\mathbf{u},\mathbf{v}] = \int_{t_1}^{t_2} \left[\mathbf{v}^2 \, \frac{\partial^2}{\partial \mathbf{u}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + \dot{\mathbf{v}}^2 \, \frac{\partial^2}{\partial \dot{\mathbf{u}}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) \right] \, \mathrm{d}t \ \ .$$

Clearly, $\mathrm{Q}[\mathbf{u},\mathbf{v}] > 0$ if and only if the integrand is positive definite. This immediately yields the desired result,

$$\mathbf{v}^2 \, \frac{\partial^2}{\partial \mathbf{u}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) + \dot{\mathbf{v}}^2 \, \frac{\partial^2}{\partial \dot{\mathbf{u}}^2} \, \mathrm{L}\left(\mathbf{u},\dot{\mathbf{u}},t\right) > 0 \ \ .$$

Since $\mathbf{v}$ is a random variation of $\mathbf{u}$, we require that this holds for every nonzero variation $\mathbf{v}$. □

Although we omit it here, it is interesting to point that if we replace the positive definiteness in Theorem 3.2 with negative definiteness everywhere we obtain a similar result for the local maxima of a functional.

Theorems 2.2 and 3.2 provide a complete framework for finding minimizing functions for any functional. We will abbreviate equations (15) using the standard notation of this text,

$$\mathbf{v}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} + \dot{\mathbf{v}}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} > 0 \ \ .$$

Example 3.3 (Planar Geodesic Revisited)**.** Recall from Example 2.3 that the shortest path between two points in the plane is the straight line given in equation (6). However, we did not actually prove that this function is a minimizer of the associated functional, so at most we know that it is a critical function. To prove that it is also a minimizer, we must look at the second variation. Since the Lagrangian is given by

$$\mathrm{L}\left(u,\dot{u},x\right) = \sqrt{1+\dot{u}^2} \ \ ,$$

we have that

$$\frac{\partial^2 \mathrm{L}}{\partial u^2} = 0 \ \ ,$$
$$\frac{\partial^2 \mathrm{L}}{\partial u \partial \dot{u}} = 0 \ \ ,$$

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{u}^2} = \frac{\partial}{\partial \dot{u}}\left(\frac{\dot{u}}{\sqrt{1+\dot{u}^2}}\right) = \frac{1}{\left(1+\dot{u}^2\right)^{3/2}} \quad .$$

For our critical function we have

$$u = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)x + \frac{y_1 x_2 - x_1 y_2}{x_2 - x_1} \quad \Rightarrow \quad \dot{u} = \frac{y_2 - y_1}{x_2 - x_1} \quad ,$$

so, defining $\alpha \equiv (y_2 - y_1) / (x_2 - x_1)$, we obtain

$$v^2 \, \frac{\partial^2 \mathrm{L}}{\partial u^2} + 2v\dot{v} \, \frac{\partial^2 \mathrm{L}}{\partial u \partial \dot{u}} + \dot{v}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \dot{u}^2} = \dot{v}^2 \left(\frac{1}{\left(1+\dot{u}^2\right)^{3/2}}\right) = \frac{\dot{v}^2}{\left(1+\alpha^2\right)^{3/2}} \quad .$$

According to Theorem 3.2, to be a minimizer we require that u satisfy $\dot{v}^2 / \left(1+\alpha^2\right)^{3/2} > 0$ for any nonzero variation v. It is clear that $\left(1+\alpha^2\right)^{3/2} > 0$, so we need only check that $\dot{v}^2 > 0$. Since v is a nonzero variation of u satisfying the boundary conditions $v\left(x_1\right) = v\left(x_2\right) = 0$, we necessarily have that v is nonconstant. This gives us that $\dot{v} \neq 0$, so $\dot{v}^2 > 0$. Thus, u satisfies Theorem 3.2 and is a minimizer of the planar arclength functional. We have finally proven that the shortest path between two points in the plane is indeed a straight line.

Unfortunately, equations (15) are often difficult to evaluate in practice. For this reason, we would like to have an alternative approach to proving positive definiteness of the second variation. Notice that if we integrate the mixed partial derivative term in equations (13) by parts we obtain an alternative form of the second variation:

$$\begin{aligned}
\mathrm{Q}[\mathbf{u},\mathbf{v}] &= \int_{t_1}^{t_2} \left[\mathbf{v}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} + 2\mathbf{v}\dot{\mathbf{v}} \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}} + \dot{\mathbf{v}}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2}\right] \, \mathrm{d}t \\
&= \int_{t_1}^{t_2} \left[\mathbf{v}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} + \dot{\mathbf{v}}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2}\right] \, \mathrm{d}t + \left(\mathbf{v}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \mathbf{v}^2 \, \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right) \, \mathrm{d}t \\
&= \int_{t_1}^{t_2} \left[\dot{\mathbf{v}}^2 \, \frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} + \mathbf{v}^2 \left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\right)\right] \, \mathrm{d}t \quad ,
\end{aligned}$$

where the third equality takes advantage of the boundary conditions on v, $\mathbf{v}\left(t_1\right) = \mathbf{v}\left(t_2\right) = 0$. If we fix the function $\mathbf{u}$ and define $\mathrm{R} \equiv \partial^2 \mathrm{L} / \partial \dot{\mathbf{u}}^2$, $\mathrm{S} \equiv \partial^2 \mathrm{L} / \partial \mathbf{u}^2 - \frac{\mathrm{d}}{\mathrm{d}t}\left(\partial^2 \mathrm{L} / \partial \mathbf{u} \partial \dot{\mathbf{u}}\right)$, then the second variation becomes a functional over the domain of admissible variations,

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(\mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2\right) \, \mathrm{d}t \quad ,$$

with Lagrangian $\mathrm{L}\left(\mathbf{v}, \dot{\mathbf{v}}, t\right) = \mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2$, which is the standard quadratic functional. Thus, in order to determine when the second variation $\mathrm{Q}[\mathbf{u},\mathbf{v}]$ is positive definite we can choose to concentrate instead on finding conditions under which the quadratic functional $\mathrm{J}[\mathbf{v}]$ is positive definite. Our treatment of this subject draws from that of [GF63].

Proposition 3.4. *Let* $\mathbf{v} : [t_1, t_2] \to \mathbb{R}^k$ *be of class* C^1 *such that* $\mathbf{v}\left(t_1\right) = \mathbf{v}\left(t_2\right) = 0$, $\mathbf{v} \neq 0$, *and let R, S be continuous functions of t. If the functional*

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(R\dot{\mathbf{v}}^2 + S\mathbf{v}^2\right) \, \mathrm{d}t$$

is positive definite, then $R \geq 0$ *for all* $t \in [t_1, t_2]$.

Proof. It will suffice to show that if $\mathrm{R} < 0$ for some $t \in [t_1, t_2]$, then $\mathrm{J}[\mathbf{v}] \leq 0$. Without loss of generality, let $t_0 \in [t_1, t_2]$ such that $\mathrm{R}(t_0) = -2\delta$, $\delta > 0$. Since R is continuous on the interval, there exists $\epsilon > 0$ such that $t_1 \leq t_0 - \epsilon < t_0 + \epsilon \leq t_2$ and $\mathrm{R} < -\delta$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$. Suppose we define $\mathbf{v}$ by

$$v_i = \begin{cases} \sin^2\left(\frac{\pi}{\epsilon}(t - t_0)\right) & \text{for } t \in [t_0 - \epsilon, t_0 + \epsilon] \\ 0 & \text{otherwise} \end{cases}$$

for each $i \in \{1, 2, ..., k\}$. Then,

$$\begin{aligned} \int_{t_1}^{t_2} \left(\mathrm{R}_i \dot{v}_i^2 + \mathrm{S}_i v_i^2\right) \mathrm{d}t &= \int_{t_0-\epsilon}^{t_0+\epsilon} \mathrm{R}_i \left[\frac{\pi}{\epsilon} \sin\left(\frac{2\pi}{\epsilon}(t - t_0)\right)\right]^2 \mathrm{d}t \\ &\quad + \int_{t_0-\epsilon}^{t_0+\epsilon} \mathrm{S}_i \sin^4\left(\frac{\pi}{\epsilon}(t - t_0)\right) \mathrm{d}t \\ &< \int_{t_0-\epsilon}^{t_0+\epsilon} \left(-\frac{\pi^2}{\epsilon^2}\,\delta\right) \mathrm{d}t + \int_{t_0-\epsilon}^{t_0+\epsilon} M_i \,\mathrm{d}t \\ &= -\frac{2\pi^2}{\epsilon}\,\delta + 2M_i\,\epsilon \end{aligned}$$

where $M_i = \max_{t_1 \leq t \leq t_2} |\mathrm{S}_i|$. Since $-\left(2\pi^2/\epsilon\right)\delta + 2M_i\,\epsilon < 0$ for sufficiently small ϵ, we have that

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(\sum_{i=1}^{k} \left(\mathrm{R}_i \dot{v}_i^2 + \mathrm{S}_i v_i^2\right)\right) \mathrm{d}t \leq \sum_{i=1}^{k} \left(-\frac{2\pi^2}{\epsilon}\,\delta + 2M_i\,\epsilon\right) < 0 \ .$$

Thus, $\mathrm{J}[\mathbf{v}] \leq 0$ when $R < 0$ for some $t \in [t_1, t_2]$, so we conclude that $\mathrm{R} \geq 0$ for all $t \in [t_1, t_2]$ whenever $\mathrm{J}[\mathbf{v}] > 0$, as desired. □

Proposition 3.4 leads us directly to our next theorem.

Theorem 3.5 (The Legendre Condition)**.** *Suppose* $\mathbf{u}$*,* $\mathbf{v}$ *and* L *satisfy the hypotheses of Theorem 3.2. If* $\mathbf{u}$ *is a local minimum of* $\mathrm{J}[\mathbf{u}]$*, then*

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} \geq 0 \tag{16}$$

for all $t \in [t_1, t_2]$.

Proof. In order to be a minimizer, $\mathbf{u}$ must satisfy $\mathrm{Q}[\mathbf{u}, \mathbf{v}] > 0$ for any nonzero variation $\mathbf{v}$. Since $\mathbf{u}$ is a fixed critical function of $\mathrm{J}[\mathbf{u}]$, we can rewrite the second variation as the functional

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left[\dot{\mathbf{v}}^2 \,\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} + \mathbf{v}^2 \left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\right)\right] \mathrm{d}t \ .$$

Using this equality, it follows immediately from Proposition 3.4 that if $\mathrm{Q}[\mathbf{u}, \mathbf{v}] > 0$, then

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} \geq 0$$

for all $t \in [t_1, t_2]$. Thus, equations (16) hold whenever $\mathbf{u}$ is a minimizer of $\mathrm{J}[\mathbf{u}]$, as desired. □

The Legendre condition provides us with a necessary condition for minimization. Although it is not sufficient for determining whether a critical function is a minimum, we can take advantage of it in certain examples to further narrow our list of candidate functions obtained via the Euler-Lagrange equations.

Example 3.6 (The Brachistochrone Revisited)**.** Recall from Example 2.6 that the critical function of the brachistochrone functional is the cycloid parametrized in equation (12). We want to show that this function is a strong candidate for minimization. Earlier we determined that the Lagrangian is given by

$$\mathrm{L}(u, \dot{u}, x) = \sqrt{\frac{1+\dot{u}^2}{2\mathrm{g}u}} \quad .$$

Thus,

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{u}^2} = \frac{\partial}{\partial \dot{u}} \left(\frac{\dot{u}}{\sqrt{2\mathrm{g}u\,(1+\dot{u}^2)}} \right) = \frac{1}{\sqrt{2\mathrm{g}u}\,(1+\dot{u}^2)^{3/2}} \quad .$$

Clearly, $(1+\dot{u}^2)^{3/2} > 0$. Also, notice that the parametrized cycloid equation, $u = (1/2)\,k^2\,(1-\cos t)$, is necessarily non-negative for all t. Thus, $\partial^2\mathrm{L}/\partial\dot{u}^2 \geq 0$ for all x. By Theorem 3.5, the cycloid still meets the necessary conditions for a minimizer. We now have stronger evidence to support our claim that the cycloid describes the path along which the bead falls in the least amount of time.

Let us now turn to the task of finding a sufficient condition for minimization. Legendre himself postulated that strengthening the condition of Proposition 3.4 to strictly positive definite would do the trick, however this is not enough. Luckily, we will only need one additional requirement. To this end, we introduce a new definition:

Definition 3.7. Let $\mathbf{v}$ satisfy the hypotheses of Proposition 3.4 and $\mathrm{J}[\mathbf{v}]$ be the quadratic functional,

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(\mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2 \right) \,\mathrm{d}t \quad .$$

Suppose further that $\mathbf{v}$ satisfies the associated Euler-Lagrange equations. It is easy to check that this yields the Jacobi equations:

$$\mathrm{S}\mathbf{v} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{R}\dot{\mathbf{v}}\right) = 0 \quad , \tag{17}$$

which are first order differential equations. A point $t_0 \neq t_1$ is said to be *conjugate* to t_1 if there exists a non-trivial solution $\bar{\mathbf{v}}$ to equations (17) such that $\bar{\mathbf{v}}\,(t_0) = \bar{\mathbf{v}}\,(t_1)$.

We are now ready to look for a sufficient condition. Our presentation of its development follows from the work of [GF63].

Proposition 3.8. *Let* $\mathbf{v}$*, R, and S satisfy the hypotheses of Proposition 3.4. Then, the quadratic functional*

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(R\dot{\mathbf{v}}^2 + S\mathbf{v}^2 \right) \,\mathrm{d}t$$

is positive definite whenever $R > 0$ *for all* $t \in [t_1, t_2]$ *and there is no point* $t_0 \in (t_1, t_2)$ *conjugate to* t_1.

Proof. Suppose there exists a function $\mathbf{w} \in C^1$ such that $\mathbf{w}^2 = \mathrm{R}(\mathrm{S} + \dot{\mathbf{w}})$ for all $t \in [t_1, t_2]$. Consider the derivative $\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{w}\mathbf{v}^2)$ of this function. Since $\mathbf{v}$ satisfies the boundary conditions $\mathbf{v}(t_1) = \mathbf{v}(t_2) = 0$, it is clear that

$$\int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{w}\mathbf{v}^2\right)\ \mathrm{d}t = 0 \ ,$$

so we have that

$$\int_{t_1}^{t_2} \left(\mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2\right)\ \mathrm{d}t = \int_{t_1}^{t_2} \left[\mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2 + \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{w}\mathbf{v}^2\right)\right]\ \mathrm{d}t \ .$$

Looking closer at the integrand, we see

$$\begin{aligned} \mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2 + \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{w}\mathbf{v}^2\right) &= \mathrm{R}\dot{\mathbf{v}}^2 + 2\mathbf{w}\mathbf{v}\dot{\mathbf{v}} + (\mathrm{S} + \dot{\mathbf{w}})\,\dot{\mathbf{v}}^2 \\ &= \mathrm{R}\dot{\mathbf{v}}^2 + 2\mathbf{w}\mathbf{v}\dot{\mathbf{v}} + \frac{\mathbf{w}^2\mathbf{v}^2}{\mathrm{R}} \\ &= \mathrm{R}\left(\dot{\mathbf{v}}^2 + \frac{2\mathbf{w}\mathbf{v}\dot{\mathbf{v}}}{\mathrm{R}} + \frac{\mathbf{w}^2\mathbf{v}^2}{\mathrm{R}^2}\right) \\ &= \mathrm{R}\left(\dot{\mathbf{v}} + \frac{\mathbf{w}\mathbf{v}}{\mathrm{R}}\right)^2 \ , \end{aligned}$$

which necessarily yields

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left(\mathrm{R}\dot{\mathbf{v}}^2 + \mathrm{S}\mathbf{v}^2\right)\ \mathrm{d}t = \int_{t_1}^{t_2} \mathrm{R}\left(\dot{\mathbf{v}} + \frac{\mathbf{w}\mathbf{v}}{\mathrm{R}}\right)^2\ \mathrm{d}t \ .$$

Thus, $\mathrm{J}[\mathbf{v}]$ is positive definite whenever $\mathrm{R} > 0$, provided the function $\mathbf{w}$ exists. Using the substitution $\mathbf{w} = -\mathrm{R}\left(\dot{\mathbf{h}}/\mathbf{h}\right)$ for some new function $\mathbf{h} \in C^1$, we obtain

$$\begin{aligned} \mathbf{w}^2 = \mathrm{R}(\mathrm{S} + \dot{\mathbf{w}}) \quad &\Rightarrow \quad \frac{\mathrm{R}^2\dot{\mathbf{h}}^2}{\mathbf{h}^2} = \mathrm{R}\left(\mathrm{S} - \frac{\left(\mathrm{R}\ddot{\mathbf{h}} + \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{R})\,\dot{\mathbf{h}}\right)\mathbf{h} - \mathrm{R}\dot{\mathbf{h}}^2}{\mathbf{h}^2}\right) \\ &\Rightarrow \quad \mathrm{S}\mathbf{h} - \left(\mathrm{R}\ddot{\mathbf{h}} + \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{R})\,\dot{\mathbf{h}}\right) = 0 \\ &\Rightarrow \quad \mathrm{S}\mathbf{h} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{R}\dot{\mathbf{h}}\right) = 0 \ , \end{aligned}$$

which are just the Jacobi equations for $\mathrm{J}[\mathbf{v}]$. Since $\mathbf{w}$ only exists for all $t \in [t_1, t_2]$ if a nonzero $\mathbf{h}$ exists, we want $\mathbf{h}$ to be a non-trivial solution to the Jacobi equations which has no conjugate points to t_1 in the interval (t_1, t_2). Therefore, $\mathrm{J}[\mathbf{v}] > 0$ whenever $\mathrm{R} > 0$ and there is no point $t_0 \in (t_1, t_2)$ conjugate to t_1, as desired. □

Proposition 3.8 finally allows us to formulate our sufficient condition for minimization, which we summarize with our next theorem.

Theorem 3.9. *Suppose* $\mathbf{u}$, $\mathbf{v}$, *and* L *satisfy the hypotheses of Theorem 3.2. Then,* $\mathbf{u}$ *is a local minimum of* $\mathrm{J}[\mathbf{u}]$ *if*

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} > 0$$

for all $t \in [t_1, t_2]$ and the Jacobi equations,

$$\left[\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\right] \mathbf{v} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2}\, \dot{\mathbf{v}}\right) = 0 \ ,$$

have only the trivial solution, $\mathbf{v} \equiv 0$, for all $t \in [t_1, t_2]$.

Proof. In order to be a minimizer, $\mathbf{u}$ must satisfy $\mathrm{Q}[\mathbf{u}, \mathbf{v}] > 0$ for any nonzero variation $\mathbf{v}$. Since $\mathbf{u}$ is a fixed critical function of $\mathrm{J}[\mathbf{u}]$, we can rewrite the second variation as the quadratic functional

$$\mathrm{J}[\mathbf{v}] = \int_{t_1}^{t_2} \left[\dot{\mathbf{v}}^2\, \frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} + \mathbf{v}^2 \left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\right)\right] \,\mathrm{d}t \ .$$

Using this equality, it follows immediately from Proposition 3.8 that $\mathrm{Q}[\mathbf{u}, \mathbf{v}] > 0$ if

$$\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2} > 0$$

for all $t \in [t_1, t_2]$ and there is no point $t_0 \in (t_1, t_2)$ conjugate to t_1. Since $\mathbf{v}(t_1) = \mathbf{v}(t_2) = 0$, the latter requirement necessitates that the Jacobi equations,

$$\left[\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u}^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \mathbf{u} \partial \dot{\mathbf{u}}}\right)\right] \mathbf{v} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 \mathrm{L}}{\partial \dot{\mathbf{u}}^2}\, \dot{\mathbf{v}}\right) = 0 \ ,$$

have only the trivial solution, $\mathbf{v} \equiv 0$, for all $t \in [t_1, t_2]$. Thus, $\mathbf{u}$ is a minimizer of $\mathrm{J}[\mathbf{u}]$ if $\partial^2 \mathrm{L}/\partial \dot{\mathbf{u}}^2$ is positive definite for all $t \in [t_1, t_2]$ and the Jacobi equations have only the trivial solution for all $t \in [t_1, t_2]$, as desired. ☐

Theorem 3.9 will prove to be a much more efficient way of determining whether or not a critical function for a functional is actually a minimizer. Again, however, we encounter an issue: finding non-trivial solutions to the Jacobi equations is often a difficult task. Fortunately, there is an analytic way of determining conjugate points acquired through detailed analysis of the Jacobi equations and applications of the works of Hilbert and Picard. Although conjugate point theory and the Jacobi equations are integral to our application of the second variation, we will not pursue them here. Instead, we state the desired result without proof. For a more complete analysis, see [Ben08].

Lemma 3.10. *Let $\mathbf{y} : [t_1, t_2]$ be of class C^2 depending on two parameters, α and β, $\mathbf{v} \in C^2$ be a variation of $\mathbf{y}$, and R and S be continuous functions of t. Consider the Jacobi equations*

$$S\mathbf{v} - \frac{\mathrm{d}}{\mathrm{d}t}\left(R\dot{\mathbf{v}}\right) = 0 \ . \tag{18}$$

We assert that there is a non-trivial solution to equations (18) for $t \in [t_1, t_2]$ whenever the determinant

$$\begin{vmatrix} \frac{\partial}{\partial \alpha}\left(\mathbf{y}_c\right) & \frac{\partial}{\partial \beta}\left(\mathbf{y}_c\right) \\ \frac{\partial}{\partial \alpha}\left(\mathbf{y}_1\right) & \frac{\partial}{\partial \beta}\left(\mathbf{y}_1\right) \end{vmatrix} \ ,$$

where $\mathbf{y}_1 = \mathbf{y}(t_1)$ and $\mathbf{y}_c = \mathbf{y}(t_c)$ for $t_c \in (t_1, t_2]$, vanishes. Any such point t_c is necessarily conjugate to t_1.

4. Some Variational Examples

We will now work through two example problems requiring the use of variational calculus in order to illustrate some of the applications of the first and second variations. These and other examples are mentioned in [Olv12].

Example 4.1 (Minimal Surface of Revolution)**.** Let (x_1, y_1), (x_2, y_2) be two points in the plane such that $x_1 < x_2$, $y_1 > 0$, and $y_2 > 0$. Moreover, let $u : [x_1, x_2] \to \mathbb{R}$ be of class C^2 satisfying the boundary conditions $u(x_1) = y_1$, $u(x_2) = y_2$ with $u \geq 0$ for $x \in (x_1, x_2)$.[1] Suppose we revolve u about the x-axis and consider the surface area of the volume formed, neglecting the circular vertical surfaces at either end. We want to determine the function u with least surface area. Thus, our goal is to minimize the area of the surface formed by revolutions of u, which is called a surface of revolution. The solution is not necessarily intuitive.

First, we must determine the surface area Lagrangian. Let S_x denote the surface area of a surface of revolution formed from revolving the function u about the x-axis. Then, we have that

$$\Delta S_x = 2\pi u \Delta S \ \ ,$$

where $\Delta S = \Delta x\sqrt{1+\dot{u}^2}$ is the arclength element we derived in Example 2.6. Explicitly, we have

$$\Delta S_x = 2\pi u \Delta x \sqrt{1+\dot{u}^2} \ \ ,$$

so the surface area Lagrangian is

$$\mathrm{L}(u, \dot{u}, x) = \frac{\Delta S_x}{\Delta x} = 2\pi u\sqrt{1+\dot{u}^2} \ \ .$$

Thus, to find the curve whose surface of revolution has minimal surface area between (x_1, y_1) and (x_2, y_2), we want to minimize the functional

$$\mathrm{J}[u] = \int_{x_1}^{x_2} 2\pi u\sqrt{1+\dot{u}^2}\,\mathrm{d}x \ \ .$$

From Theorem 2.2, we require that any minimizer of $\mathrm{J}[u]$ satisfy the Euler-Lagrange Equations. By Theorem 2.5, any minimizer must also satisfy the equivalent Hamilton's Equations. Since $\partial \mathrm{L}/\partial x = 0$, Proposition 2.7 tells us that the Hamiltonian is a first integral solution. Using Hamilton's Equations, we can write the Hamiltonian as

$$\mathrm{H} = p\dot{u} - \mathrm{L} = \dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} \ \ ,$$

from which we obtain

$$\dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} = C$$

for some constant C. Thus, we see that

$$\begin{aligned} &\dot{u}\frac{\partial \mathrm{L}}{\partial \dot{u}} - \mathrm{L} = C \\ \Rightarrow \quad &\dot{u}\left(\frac{\partial}{\partial \dot{u}}\left(2\pi u\sqrt{1+\dot{u}^2}\right)\right) - 2\pi u\sqrt{1+\dot{u}^2} = C \end{aligned}$$

[1]The requirement that u be nonnegative is strictly for simplicity in the calculations that follow. Given any curve in the plane, we can always revolve it about an arbitrary horizontal axis as long as the curve does not cross that axis.

$$\Rightarrow \quad \frac{2\pi u\dot{u}^2}{\sqrt{1+\dot{u}^2}} - 2\pi u\sqrt{1+\dot{u}^2} = C$$
$$\Rightarrow \quad 2\pi\left(\frac{-u}{\sqrt{1+\dot{u}^2}}\right) = C \ .$$

This leads us to the conclusion that any minimizing curve must satisfy the differential equation

$$\frac{u}{\sqrt{1+\dot{u}^2}} = k\ , \qquad \text{where} \ \ k \equiv \frac{-C}{2\pi} \ .$$

This equation is separable and can be solved by direct integration:

$$\dot{u} = \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{\sqrt{u^2-k^2}}{k}$$
$$\Rightarrow \quad \int \frac{k}{\sqrt{u^2-k^2}}\,\mathrm{d}u = \int \mathrm{d}x$$
$$\Rightarrow \quad k\cosh^{-1}\left(\frac{u}{k}\right) = x + C$$
$$\Rightarrow \quad u = k\cosh\left(\frac{x+C}{k}\right)$$

where C is now a constant of integration. From this result, we see that the critical function is given by

$$u = a\cosh\left(\frac{x-b}{a}\right) \ , \tag{19}$$

for some constants a and b. This curve is called a catenary and its surface of revolution is known as a catenoid. The constants a and b are determined from the boundary conditions,

$$y_1 = a\cosh\left(\frac{x_1-b}{a}\right) \ ,$$
$$y_2 = a\cosh\left(\frac{x_2-b}{a}\right) \ .$$

All that remains to be proven is that the catenoid is the minimal surface of revolution. Theorem 3.9 says this can be done by showing that $\partial^2\mathrm{L}/\partial\dot{u}^2 > 0$ and the Jacobi equation,

$$\left[\frac{\partial^2\mathrm{L}}{\partial u^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2\mathrm{L}}{\partial u\partial\dot{u}}\right)\right]v - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2\mathrm{L}}{\partial\dot{u}^2}\,\dot{v}\right) = 0 \ ,$$

has only the trivial solution, $v \equiv 0$, for all $x \in [x_1, x_2]$. Notice that

$$\frac{\partial^2\mathrm{L}}{\partial\dot{u}^2} = \frac{\partial}{\partial\dot{u}}\left(2\pi u\sqrt{1+\dot{u}^2}\right) = \frac{u}{\left(1+\dot{u}^2\right)^{3/2}} > 0$$

since $u = a\cosh\left(\frac{x-b}{a}\right)$ is always postive, provided $a > 0$, so we need only show that the Jacobi equation has no nontrivial solution in the interval. According to Lemma 3.10, we can do this by examing solutions of the equation

$$\begin{vmatrix} \frac{\partial}{\partial a}(u_c) & \frac{\partial}{\partial b}(u_c) \\ \frac{\partial}{\partial a}(u_1) & \frac{\partial}{\partial b}(u_1) \end{vmatrix} = 0 \ ,$$

where $u_1 = u(x_1)$ and $u_c = u(x_c)$ for some $x_c \in (x_1, x_2]$. Noting that

$$\frac{\partial u}{\partial a} = \cosh\left(\frac{x-b}{a}\right) - \left(\frac{x-b}{a}\right)\sinh\left(\frac{x-b}{a}\right)$$
$$\frac{\partial u}{\partial b} = -\sinh\left(\frac{x-b}{a}\right) \ ,$$

expanding this determinant gives us

$$\begin{vmatrix} \cosh(z_c) - (z_c)\sinh(z_c) & -\sinh(z_c) \\ \cosh(z_1) - (z_1)\sinh(z_1) & -\sinh(z_1) \end{vmatrix}$$
$$= \sinh(z_c)\cosh(z_1) - \cosh(z_c)\sinh(z_1) + [(z_c) - (z_1)]\sinh(z_c)\sinh(z_1) \ ,$$

where we have substituted $z_1 = (x_1 - b)/a$ and $z_c = (x_c - b)/a$ for simplicity. Clearly, the determinant is zero when $z_c = z_1$, which requires that $x_c = x_1$. However, x_1 does not fall in our interval of concern. If we assume $x_1 \neq b$ (i.e. $z_1 \neq 0$) and solve for points where the determinant vanishes we obtain the equation

$$\coth\left(\frac{x_c-b}{a}\right) - \left(\frac{x_c-b}{a}\right) = \coth\left(\frac{x_1-b}{a}\right) - \left(\frac{x_1-b}{a}\right) \ ,$$

which has one solution other than $x_c = x_1$. It turns out that this solution is within the interval $(x_1, x_2]$ under certain circumstances, but in order to discuss this we must understand a little more about our catenary curve. For the two points, (x_1, y_1) and (x_2, y_2), there are generally two catenaries which pass through both points. For the sake of simplicity, let us assume that $y_1 < y_2$ (a similar argument exists for the case where $y_1 > y_2$). Then, y_1 lies on the descending branch of one of these catenaries and the ascending branch of the other, while y_2 lies on the ascending branch of both. These two catenaries are called the *deep* and *shallow* catenaries connecting y_1 and y_2, respectively. As discussed in [Bol61], the conjugate point x_c may lie in the interval for the deep catenary but not for the shallow catenary. Unfortunately, the surface area of the shallow catenary is greater than that of the deep catenary, so it is usually not a viable minimizing curve. If the difference $|y_2 - y_1|$ is not too small relative to $|x_2 - x_1|$, the conjugate point of the deep catenary may fall outside of our interval of concern, in which case it is the minimizing curve. If the difference $|y_2 - y_1|$ is too small relative to $|x_2 - x_1|$, however, the conjugate point of the deep catenary will fall within the interval. In this case, the catenary is not the minimal surface of revolution. It can be shown that the Goldschmidt discontinuous solution has minimal surface area under these conditions, but we will not explore this here. In the end we conclude that, provided $|y_2 - y_1|$ is not too small relative to $|x_2 - x_1|$, the function whose surface of revolution between the points (x_1, y_1), (x_2, y_2) has minimal surface area is indeed a catenoid, given by revolutions of the catenary in equation (19).

The minimal surface of revolution is more familiar to us than we may expect. Suppose we have two thin rings of the same diameter held parallel to each other such that the separation distance between them is zero. If we dip the rings in a liquid soap solution and slowly remove them, we should see a thin film of soap stretched throughout the interior of the rings. Now, if we slowly increase the separation distance between the rings the film of soap will remain connected to both rings in such a way that the total energy of the soap molecules is minimized. It turns out that the natural shape formed by the soap film is a catenoid, the minimal surface of revolution.[2] When the

[2] In actuality, the surface is not a true catenoid due to gravity, which deforms it. However, this deformation is slight due to the relatively small mass of the soap film, so the surface appears to be a catenoid.

rings are separated by too great a distance, the soap film becomes unstable and breaks, forming two separate surfaces in the interior of each ring. This configuration is that of the Goldschmidt discontinuous solution.

Example 4.2 (Spherical Geodesic)**.** Let (r_1, θ_1, ϕ_1), (r_2, θ_2, ϕ_2) be two points lying on a sphere of radius R in spherical coordinates. Moreover, let the functions $r : [0, t_1] \to \mathbb{R}^{\geq 0}$, $\theta : [0, t_1] \to [0, \pi]$, and $\phi : [0, t_1] \to [0, 2\pi)$ represent a parametrization of the curve $\mathbf{u} : (r, \theta, \phi) \to \mathbb{R}^3$ satisfying the boundary conditions $r\,(t) = R$ for all t, $\theta\,(0) = 0$, $\theta\,(t_1) = \theta_0$, and $\phi\,(t_1) = \phi_0$ (i.e. $\mathbf{u}\,(r_1, \theta_1, \phi_1) = (R, 0, \cdot)$ and $\mathbf{u}\,(r_2, \theta_2, \phi_2) = (R, \theta_0, \phi_0)$).[3] Clearly, $\mathbf{u}$ is bounded to the sphere of radius R, and we assume that $\theta(t)$ and $\phi(t)$ are both of class C^2. We want to determine the function $\mathbf{u}$ which describes the minimum distance between two points on the sphere. Thus, our goal is to minimize the arclength, confined to the surface of the sphere, between the points (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2).

First, we must determine the arclength Lagrangian for curves confined to the surface of the sphere. To do so, consider the three dimensional Cartesian arclength element,

$$\Delta S = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}\ . \tag{20}$$

To constrain ΔS to the surface of the sphere, we will use the spherical coordinate substitution

$$\begin{aligned} x &= R\sin(\theta)\cos(\phi) \\ y &= R\sin(\theta)\sin(\phi) \\ z &= R\cos(\theta)\ \ , \end{aligned}$$

from which we obtain

$$\begin{aligned} (\Delta x)^2 =& R^2\cos^2(\theta)\cos^2(\phi)\,(\Delta\theta)^2 + R^2\sin^2(\theta)\sin^2(\phi)\,(\Delta\phi)^2 \\ &- 2R^2\sin(\theta)\cos(\theta)\sin(\phi)\cos(\phi)\Delta\theta\Delta\phi \\ (\Delta y)^2 =& R^2\cos^2(\theta)\sin^2(\phi)\,(\Delta\theta)^2 + R^2\sin^2(\theta)\cos^2(\phi)\,(\Delta\phi)^2 \\ &+ 2R^2\sin(\theta)\cos(\theta)\sin(\phi)\cos(\phi)\Delta\theta\Delta\phi \\ (\Delta z)^2 =& R^2\sin^2(\theta)\Delta\theta^2 \qquad\qquad\qquad\qquad\qquad\qquad . \end{aligned}$$

Then, the squared spherical arclength element is

$$\begin{aligned} (\Delta S)^2 =& R^2\left[\cos^2(\theta)\cos^2(\phi) + \cos^2(\theta)\sin^2(\phi) + \sin^2\theta\right](\Delta\theta)^2 \\ &+ R^2\left[\sin^2(\theta)\sin^2(\phi) + \sin^2(\theta)\cos^2(\phi)\right](\Delta\phi)^2 \qquad , \end{aligned}$$

which can be simplified to

$$\Delta S = R\sqrt{(\Delta\theta)^2 + \sin^2(\theta)\,(\Delta\phi)^2}\ . \tag{21}$$

Evidently, the spherical arclength Lagrangian is

$$\mathrm{L}(\theta, \phi, \dot{\theta}, \dot{\phi}, t) = R\sqrt{\dot{\theta}^2 + \dot{\phi}^2\sin^2(\theta)}\ \ ,$$

[3]Using the point $(R, 0, \cdot)$ as our initial point is done without loss of generality to make computation easier. The result can be extended to any inititial point on the sphere.

so finding the curve on the surface of the sphere with minimal arclength between (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) requires minimizing the functional

$$\mathrm{J}[\theta, \phi] = \int_0^{t_1} R\sqrt{\dot{\theta}^2 + \dot{\phi}^2 \sin^2(\theta)}\ \mathrm{d}t \ \ .$$

From Theorem 2.2, we require that any minimizer of $\mathrm{J}[\theta, \phi]$ satisfy the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\theta}}\right) = \frac{\partial \mathrm{L}}{\partial \theta} \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\phi}}\right) = \frac{\partial \mathrm{L}}{\partial \phi} \ \ .$$

The Euler-Lagrange equation for θ does not reveal much about the desired curve. However, that for ϕ is very useful, since $\partial \mathrm{L}/\partial \phi = 0$. This gives us

$$\frac{\partial \mathrm{L}}{\partial \phi} = 0 \ \ \Rightarrow \ \ \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\phi}}\right) = 0 \ \ \Rightarrow \ \ \frac{\partial \mathrm{L}}{\partial \dot{\phi}} = C \ \ ,$$

for some constant C. Thus, we have

$$\begin{aligned} &\frac{\partial \mathrm{L}}{\partial \dot{\phi}} = C \\ \Rightarrow \ \ &\frac{\partial}{\partial \dot{\phi}}\left(R\sqrt{\dot{\theta}^2 + \dot{\phi}^2 \sin^2(\theta)}\right) = C \\ \Rightarrow \ \ &\frac{R\dot{\phi}\sin^2(\theta)}{\sqrt{\dot{\theta}^2 + \dot{\phi}^2 \sin^2(\theta)}} = C \ \ . \end{aligned}$$

Since this must be true for all possible values of θ, we choose $\theta = 0$ in order to solve for C:

$$\theta = 0 \ \ \Rightarrow \ \ C = 0 \, .$$

This tells us that, for all possible θ, we have

$$\begin{aligned} &\dot{\phi}\sin^2(\theta) = 0 \\ \Rightarrow \ \ &\dot{\phi} = 0 \\ \Rightarrow \ \ &\phi = k \ \ , \end{aligned}$$

for some constant k. The boundary condition $\phi(t_1) = \phi_0$ further tells us that $k = \phi_0$. Thus, the curve $\mathbf{u}$ which minimizes the distance between the points (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) is parametrized by functions θ and ϕ such that

$$\theta = g(t) \qquad \text{and} \qquad \phi = \phi_0 \ \ ,$$

for some increasing function $g : [0, t_1] \to [0, \pi]$. This means that $\mathbf{u}$ is bound to an arc of the great circle defined by $\phi = \phi_0$.

We must now determine the equation of the great circle arc connecting our two points. To do so, note that

$$\begin{aligned} x\sin(\phi_0) &= [R\cos(\phi_0)\sin(\theta)]\sin(\phi_0) \\ y\cos(\phi_0) &= [R\sin(\phi_0)\sin(\theta)]\cos(\phi_0) \ \ , \end{aligned}$$

which means we have

$$x\sin(\phi_0) - y\cos(\phi_0) = 0 \ \ .$$

Thus, the great circle can be expressed as the set of points

$$\left\{(x,y,z)\in\mathbb{R}^3 \mid x\sin(\phi_0)-y\cos(\phi_0)=0\ ,\ x^2+y^2+z^2=R^2\right\}$$

and the arc on which the curve $\mathbf{u}$ lies can be parametrized by

$$\mathbf{u}(x,y,z)=R\left(\cos(\phi_0)\sin(t),\sin(\phi_0)\sin(t),\cos(t)\right)\quad,\ t\in[0,\pi]\quad. \tag{22}$$

The curve $\mathbf{u}$ is necessarily a critical function of the spherical arclength functional. All that remains to be proven is that equation (22), with boundary conditions $(r,\theta,\phi)\,|_{t=0}=(R,0,\phi_0)$, $(r,\theta,\phi)\,|_{t=t_1}=(R,\theta_0,\phi_0)$, is actually the minimizing curve. By Theorem 3.9, we must show that $\partial^2\mathrm{L}/\partial\dot{\phi}^2>0$ and the Jacobi equation,

$$\left[\frac{\partial^2\mathrm{L}}{\partial\phi^2}-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2\mathrm{L}}{\partial\phi\partial\dot{\phi}}\right)\right]v-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2\mathrm{L}}{\partial\dot{\phi}^2}\,\dot{v}\right)=0\ ,$$

has only the trivial solution, $v\equiv 0$, for all $\theta\in[0,\pi]$. We see fairly easily that $\partial^2\mathrm{L}/\partial\dot{\phi}^2>0$ in the interval, since $\dot{\theta}>0$ implies

$$\frac{\partial^2\mathrm{L}}{\partial\dot{\phi}^2}=\frac{\partial}{\partial\dot{\phi}}\left(R\sqrt{\dot{\theta}^2+\dot{\phi}^2\sin^2(\theta)}\right)=\frac{R\dot{\phi}\sin^2(\theta)}{\sqrt{\dot{\theta}^2+\dot{\phi}^2\sin^2(\theta)}}>0\ .$$

However, determining solutions to the Jacobi equation is difficult in this notation. For this, we will utilize the fact that any great circle lies entirely in one plane.

Without loss of generality, assume $\mathbf{u}$ lies entirely in the x-y plane, so that we can write the curve completely in terms of x and y. Thus, we have $x^2+y^2=R^2$. Recall that we also have $x=R\cos(\phi_0)\sin(\theta)$, so we can express the curve as a function of θ:

$$y=R\sqrt{1-\cos^2(\phi_0)\sin^2(\theta)}\ ,\ \theta\in[0,\pi]\ .$$

We now have a two parameter family of curves in the plane, which we will use in order to apply Lemma 3.10. According to the lemma, we can find conjugate points by examing solutions of the equation

$$\begin{vmatrix}\frac{\partial}{\partial\phi_0}(y_c) & \frac{\partial}{\partial R}(y_c)\\ \frac{\partial}{\partial\phi_0}(y_1) & \frac{\partial}{\partial R}(y_1)\end{vmatrix}=0\ ,$$

where $y_1=y(\theta_1)$ and $y_c=y(\theta_c)$ for some $\theta_c\in(0,\pi]$. Noting that

$$\frac{\partial y}{\partial\phi_0}=\frac{R\sin(\phi_0)\cos(\phi_0)\sin^2(\theta)}{\sqrt{1-\cos^2(\phi_0)\sin^2(\theta)}}$$

$$\frac{\partial y}{\partial R}=\sqrt{1-\cos^2(\phi_0)\sin^2(\theta)}\ ,$$

expanding the determinant gives us

$$\begin{vmatrix}\frac{R\sin(\phi_0)\cos(\phi_0)\sin^2(\theta_c)}{\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_c)}} & \sqrt{1-\cos^2(\phi_0)\sin^2(\theta_c)}\\ \frac{R\sin(\phi_0)\cos(\phi_0)\sin^2(\theta_1)}{\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_1)}} & \sqrt{1-\cos^2(\phi_0)\sin^2(\theta_1)}\end{vmatrix}$$

$$=\frac{R\sin(\phi_0)\cos(\phi_0)\sin^2(\theta_c)\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_1)}}{\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_c)}}$$

$$- \frac{R\sin(\phi_0)\cos(\phi_0)\sin^2(\theta_1)\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_c)}}{\sqrt{1-\cos^2(\phi_0)\sin^2(\theta_1)}} \quad .$$

After some manipulation, we find that the determinant is zero when $\sin^2(\theta_c) = \sin^2(\theta_1)$, which has solutions $\theta_c = \theta_1 + \pi k$, $k \in \mathbb{Z}$. The only one of these in the interval we are concerned about, $(0, \pi]$, is $\theta_c = \pi$. Thus, the curve $\mathbf{u}$ only contains a conjugate point when the points $(R, 0, \phi_0)$ and (R, θ_0, ϕ_0) are antipodal. In the antipodal case, any great circle arc between the two points has the same length, so there is no single minimizing curve. Therefore, the function describing the shortest path between two points on a sphere of radius R which are not antipodes is indeed an arc of the great circle connecting them, parametrized by equation (22).

5. Lagrangian Mechanics

The theory of classical mechanics involves using Newton's equations of dynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{3n} \mathrm{m}\dot{r}_i\right) = -\nabla U \quad ,$$

for a given mechanical system, where U, m and $\mathbf{r}$ describe the potential energy, mass, and position, respectively, of a system of n physical objects, in order to determine equations governing the motion of the objects. Lagrange successfully applied the principles of the calculus of variations directly to the theory of classical mechanics, which Hamilton later refined. The result was what became known as the Lagrangian method, or, more generally, Lagrangian mechanics. Our goal here is to develop the theory of Lagrangian mechanics and illustrate some of its applications. We begin with a proposition:

Proposition 5.1. *Let* $\mathbf{r} : [t_1, t_2] \to \mathbb{R}^{3k}$ *be of class* C^2 *satisfying Newton's equations of dynamics with mass* m*, potential energy* $U = U(\mathbf{r})$*, and kinetic energy* $T = \sum_{i=1}^{3k} \left(\mathrm{m}\dot{r}_i{}^2/2\right)$*. Then,* $\mathbf{r}$ *is a critical function of the functional*

$$\mathrm{J}[\mathbf{r}] = \int_{t_0}^{t_1} \mathrm{L}(\mathbf{r}, \dot{\mathbf{r}}, t)\ \mathrm{d}t \quad ,$$

where the Lagrangian, $\mathrm{L}(\mathbf{r}, \dot{\mathbf{r}}, t) = T - U$*, is the difference between the kinetic and potential energy of the system.*

Proof. A function $\mathbf{r}$ is a critical function of $\mathrm{J}[\mathbf{r}]$ if it satisfies the Euler-Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}}\right) = \frac{\partial \mathrm{L}}{\partial \mathbf{r}} \quad .$$

Using $\mathrm{L} = T - U$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial}{\partial \dot{\mathbf{r}}}(T-U)\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{3k} \mathrm{m}\dot{r}_i\right)$$
$$\frac{\partial \mathrm{L}}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}}(T-U) = -\frac{\partial U}{\partial \mathbf{r}} = -\nabla U \quad .$$

Thus, $\mathbf{r}$ satisfies the Euler-Lagrange equations whenever it satisfies Newton's equations of dynamics, $\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{3k} \mathrm{m}\dot{r}_i\right) = -\nabla U$, as desired. □

Because each $\mathbf{r}$ refers to a mechanical system, we use convenient terminology for most quantities in Lagrangian mechanics. The r_i are called *generalized coordinates*, with *generalized velocities*, $\dot{r}_i$, *generalized momenta*, $p_i = \partial \mathrm{L}/\partial \dot{r}_i$, and *generalized forces*, $\partial \mathrm{L}/\partial r_i$. The functional $\mathrm{J}[\mathbf{r}]$ is refered to as the *action* and the Euler-Lagrange equations $\frac{\mathrm{d}}{\mathrm{d}t}\left(\partial \mathrm{L}/\partial \dot{\mathbf{r}}\right) = \partial \mathrm{L}/\partial \mathbf{r}$ are called *Lagrange's equations*. Our next proposition, which builds upon Proposition 5.1 using this physical terminology, formally defines the underlying principle upon which Lagrangian mechanics is built.

Proposition 5.2 (Hamilton's Principle of Stationary Action)**.** *Let* $\mathbf{r} : [t_1, t_2] \to \mathbb{R}^{3k}$ *be of class* C^2 *and* $\mathrm{L} = T - U$ *be the Lagrangian for a mechanical system of* k *masses with kinetic and potential energy* T *and* U*, respectively. Then, each* r_i *is a generalized coordinate for one of the* k *masses if and only if the action is stationary. That is, if and only if* $\nabla \mathrm{J}[\mathbf{r}] = 0$.

Proof. If the r_i are the generalized coordinates of k masses in a physical system, then $\mathbf{r}$ necessarily satisfies Newton's equations of dynamics. Proposition 5.1 then states that $\mathbf{r}$ must also be a critical function of the action, $\mathrm{J}[\mathbf{r}]$. Thus, $\mathrm{J}[\mathbf{r}] = 0$, so the action is stationary. Conversely, if the action is stationary we must have $\mathrm{J}[\mathbf{r}] = 0$. Then, $\mathbf{r}$ necessarily satisfies Lagrange's equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}}\right) = \frac{\partial \mathrm{L}}{\partial \mathbf{r}} \quad .$$

Since $\mathrm{L} = T - U$, we can utilize the proof of Proposition 5.1 to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}}\right) - \frac{\partial \mathrm{L}}{\partial \mathbf{r}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{3k} \mathrm{m}\dot{r}_i\right) - (-\nabla U) = 0$$

$$\Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{3k} \mathrm{m}\dot{r}_i\right) = -\nabla U \quad ,$$

which are precisely Netwon's equations of dynamics. Thus, the r_i are generalized coordinates of k masses in a physical system. We conclude that the r_i are the generalized coordinates of k masses in a mechanical system if and only if the action is stationary, as desired. □

Notice that Hamilton's Principle of Stationary Action only requires that the action be stationary at $\mathbf{r}$ rather than a minimum. This means that it is not necessary to utilize the second variation when determining equations of motion via the Lagrangian method. As it turns out, equations of motion often do correspond to a local minimum of the action. For this reason, Proposition 5.2 is sometimes refered to as Hamilton's Principle of Least Action, which is somewhat of a misnomer. We will refrain from refering to the theorem by this name to avoid confusion.

Example 5.3 (Hooke's Law)**.** Suppose a mass m on a frictionless surface is attached to a horizontal spring with spring contant k. If the mass is set in motion along the axis through the center of the spring, then Newton's equations of dynamics reduce to Hooke's Law, which describes the motion of the mass in terms of the spring constant:

$$\mathrm{m}\ddot{r} = -\mathrm{k}r \quad .$$

Our goal is to verify Hooke's Law using Hamilton's Principle of Least Action. Let $r : [0, t] \to \mathbb{R}$, of class C^2, represent the position of the mass with repsect to its equilibrium positon, the point where the spring is neither stretched nor compressed. The kinetic and potential energy of the mass are $T = \mathrm{m}\dot{r}^2/2$ and $U = \mathrm{k}r^2/2$, respectively. Thus, we have that

$$\mathrm{L} = T - U = \frac{1}{2}\left(\mathrm{m}\dot{r}^2 - \mathrm{k}r^2\right) \quad .$$

By Proposition 5.2, r must cause the action to be stationary. This occurs when r satisfies the Lagrange's equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{r}}\right) = \frac{\partial \mathrm{L}}{\partial r} \quad .$$

For the spring-mass system, this yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{r}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{m}\dot{r}\right) = \mathrm{m}\ddot{r}$$

$$\frac{\partial \mathrm{L}}{\partial r} = -\mathrm{k}r \quad ,$$

so we have indeed arrived at Hooke's Law, $\mathrm{m}\ddot{r} = -\mathrm{k}r$.

Determining the equations of motion for a mechanical system via Hamilton's Principle of Stationary Action is summarized conveniently in our next theorem. The process we outline is known as the Lagrangian method.

Theorem 5.4 (The Lagrangian Method)**.** *Let* $\mathbf{r}$ *and* L *satisfy the hypotheses of Proposition 5.2. Then, Newton's equations of dynamics for a mechanical system of* k *masses are given by Lagrange's equations,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}}\right) = \frac{\partial \mathrm{L}}{\partial \mathbf{r}} \quad .$$

Proof. Since $\mathbf{r}$ and L satisfy the hypotheses of Proposition 5.2, the r_i are the generalized coordinates of the k masses if and only if the action is stationary. This occurs whenever $\mathbf{r}$ satisties Lagrange's equations. Thus, by Proposition 5.1, Lagrange's equations must be equivalent to Newton's equations of dynamics for the system. Therefore, Newton's equations of dynamics for a mechanical system of k masses are given by Lagrange's equations, as desired. □

Since Theorem 5.4 does not specifically introduce anything new, it is mostly a formality. However, it succinctly defines the Lagrangian method, reducing the steps necessary to obtain equations of motion for a mechanical system from the generalized coordinates of the masses within the system. Lagrangian mechanics is simply the utilization of the Lagrangian method to solve problems in classical mechanics.

Example 5.5 (Simple Pendulum)**.** Suppose a mass m is suspended from a massless rod of length l which hangs from a hinge allowing the rod to swing freely in a vertical plane. If this pendulum is set in motion, the mass will oscillate about its equilibrium position directly below the hinge with a predictable frequency, provided the oscillations are small. Our goal is to determine this frequency of oscillation. Let $r : [0, t] \to \mathbb{R}$ and $\theta : [0, t] \to [-\pi/2, \pi/2]$ in C^2 be the radial and angular positions of the mass with respect to the hinge, respectively. Then, the kinetic energy of the mass is

$$T = \frac{1}{2}\,\mathrm{m}\left(\dot{r}^2 + r^2\dot{\theta}^2\right)$$

and its potential energy is

$$U = \mathrm{m}\mathrm{g}l\left(1 - \cos\left(\theta\right)\right) \quad .$$

Since $r = l$ and $\dot{r} = 0$, we have

$$\mathrm{L} = T - U = \frac{1}{2}\,\mathrm{m}l^2\dot{\theta}^2 - \mathrm{mg}l\left(1 - \cos\left(\theta\right)\right) \ .$$

By Theorem 5.4, we can obtain the equation of motion for the mass from Lagrange's equations for r and θ. Since r does not enter these equations, we need only use that for θ. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\theta}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{m}l^2\dot{\theta}\right) = \mathrm{m}l^2\ddot{\theta}$$

$$\frac{\partial \mathrm{L}}{\partial \theta} = -\mathrm{mg}l\sin\left(\theta\right) \ ,$$

so we obtain the equation of motion

$$\ddot{\theta} + \frac{\mathrm{g}}{l}\sin\left(\theta\right) = 0 \ .$$

For small oscillations we have $\sin\left(\theta\right) \approx \theta$. Substituting this approximation into the equation of motion yields

$$\ddot{\theta} + \frac{\mathrm{g}}{l}\,\theta = 0 \ ,$$

which has solution $\theta = k_1\cos\left(\omega t\right) + k_2\sin\left(\omega t\right)$, where $\omega^2 = \mathrm{g}/l$ and k_1, k_2 are constants determined by initial conditions. Thus, the frequency of small oscillations is given by the well known quantity $\omega = \sqrt{\mathrm{g}/l}$.

Example 5.6 (Projectile Motion)**.** Suppose a mass m is tossed into the air at an angle and allowed to fall freely to the ground. If we neglect air resistance, the only force acting on the mass is gravity in the vertical direction. We want to determine the equations of motion for the mass while it is in the air. Let $\mathbf{r} : [0, t] \to \mathbb{R}^3$ be of class C^2 where $\mathbf{r}\left(t\right) = \left(x\left(t\right), y\left(t\right), z\left(t\right)\right)$ describes the orthogonal components of the position of the mass, with $\hat{z}$ the vertical direction. Then, the kinetic and potential energy of the mass are given by $T = \mathrm{m}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right)/2$ and $U = \mathrm{mg}z$, respectively. Thus, we have that

$$\mathrm{L} = T - U = \frac{1}{2}\,\mathrm{m}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - \mathrm{mg}z \ .$$

Theorem 5.4 states that we can obtain the equations of motion from Lagrange's equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{x}}\right) = \frac{\partial \mathrm{L}}{\partial x} \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{y}}\right) = \frac{\partial \mathrm{L}}{\partial y} \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{z}}\right) = \frac{\partial \mathrm{L}}{\partial z} \ ,$$

so we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{m}\dot{x}\right) = 0 \quad \Rightarrow \quad \mathrm{m}\dot{x} = p_x$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{m}\dot{y}\right) = 0 \quad \Rightarrow \quad \mathrm{m}\dot{y} = p_y$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{m}\dot{z}\right) = -\mathrm{mg} \quad \Rightarrow \quad \ddot{z} = -g \ ,$$

where the horizontal components of the linear momentum, p_x and p_y, are constants. This gives us that the motion of the mass is described by the functions $x = \dot{x}_0 t + x_0$, $y = \dot{y}_0 t + y_0$, $z = \dot{z}_0 t - \mathrm{g}t^2/2 + z_0$, where $\mathbf{r}\left(0\right) = \left(x_0, y_0, z_0\right)$ and $\dot{\mathbf{r}}\left(0\right) = \left(\dot{x}_0, \dot{y}_0, \dot{z}_0\right)$ are the intial position and velocity of the mass, respectively. As we expect from experience, the horizontal motion of the mass is unaffected by the force of gravity and the horizontal component of its momentum is unchanged.

Notice that the x and y independence of the projectile motion Lagrangian resulted in a conserved quantity in each case. In Lagrangian mechanics, this common result is related to mechanical conservation laws. To this end, we introduce the following definition.

Definition 5.7. Let $\mathbf{r} : \mathbb{R} \to \mathbb{R}^k$ be of class C^2 and $\mathrm{L}\left(\mathbf{r}, \dot{\mathbf{r}}, t\right)$ a Lagrangian. Then, any generalized coordinate of the system, r_i, such that $\partial \mathrm{L}/\partial r_i = 0$ is called a *cyclic coordinate*.

Cyclic coordinates correspond directly to conserved quantities. It follows from the definition of generalized momentum, $p_i = \partial \mathrm{L}/\partial \dot{r}_i$, and Lagrange's equations that the generalized momentum p_i corresponding to a cyclic coordinate r_i is conserved. Example 5.6 illustrates this concept through the conservation of linear momentum. However, generalized momenta need not be linear momenta, as our next example shows.

Example 5.8 (Projectile Motion Revisited)**.** Suppose we have the same scenario we had in Example 5.6 where a mass m is tossed into the air and allowed to fall freely to ground. This time, however, let $\mathbf{r} : [0, t] \to \mathbb{R} \times [0, \pi] \times [0, 2\pi)$ be of class C^2 where $\mathbf{r}\left(t\right) = \left(r\left(t\right), \theta\left(t\right), \phi\left(t\right)\right)$ describes the radial, polar, and azimuthal components of the position of the mass, respectively, with $\theta = 0$ the vertical direction. Then, the kinetic and potential energy of the mass are given by $T = \mathrm{m}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\left(\theta\right)\dot{\phi}^2\right)/2$ and $U = \mathrm{m g} r \cos\left(\theta\right)$, respectively. Thus, we have that

$$\mathrm{L} = T - U = \frac{1}{2}\mathrm{m}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\left(\theta\right)\dot{\phi}^2\right) - \mathrm{m g} r\cos\left(\theta\right) \ .$$

Theorem 5.4 states that we can obtain the equations of motion from Lagrange's equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{r}}\right) = \frac{\partial \mathrm{L}}{\partial r} \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\theta}}\right) = \frac{\partial \mathrm{L}}{\partial \theta} \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{L}}{\partial \dot{\phi}}\right) = \frac{\partial \mathrm{L}}{\partial \phi} \ ,$$

so we have

$$\begin{aligned}
\ddot{r} &= r\dot{\theta}^2 + r\sin^2\left(\theta\right)\dot{\phi}^2 - \mathrm{g}\cos\left(\theta\right) \\
\ddot{\theta} &= \sin\left(\theta\right)\cos\left(\theta\right)\dot{\phi}^2 - \frac{\mathrm{g}}{r}\cos\left(\theta\right) \\
\frac{\mathrm{d}}{\mathrm{d}t}&\left(\mathrm{m}r^2\sin^2\left(\theta\right)\dot{\phi}\right) = 0 \quad \Rightarrow \quad \mathrm{m}r^2\sin^2\left(\theta\right)\dot{\phi} = L_\phi \ ,
\end{aligned}$$

where the azimuthal component of the angular momentum, L_ϕ, is conserved. Although these equations for the motion of the mass are more complex than those found in Example 5.6, we obtain the same result: the horizontal motion of the mass is unaffected by the force of gravity and the horizontal component of its momentum is unchanged.

We can further apply the methods of the calculus of variations to classical mechanics by investigating the Hamiltonian. Recall that the Legendre transform of the Lagrangian, which utilizes Poisson's variables $\mathbf{p} = \partial \mathrm{L}/\partial \dot{\mathbf{u}}$, yields the equivalent Hamiltonian, $\mathrm{H} = \mathbf{p}\dot{\mathbf{u}} - \mathrm{L}$. Applying this to Newton's equations of dynamics gives us

$$\mathbf{p} = \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}} \ , \quad \mathrm{H} = \mathbf{p}\dot{\mathbf{r}} - \mathrm{L} \ ,$$

where $\mathrm{L} = T - U$. From this, we obtain the following result.

Proposition 5.9. *Let* $\mathbf{r}$*,* U*,* T*, and* L *satisfy the hypotheses of Proposition 5.1 for mass* m*. Then, the Hamiltonian is given by the total energy of the system,*

$$\mathrm{H} = T + U \ .$$

Proof. For mass m, the kinetic and potential energy are given by $T = \sum_{i=1}^{3k} (\mathrm{m}\dot{r}_i^2/2)$ and $U = U(\mathbf{r})$, respectively. Thus, we have

$$\mathbf{p} = \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{r}}} = \sum_{i=1}^{3k} (\mathrm{m}\dot{r}_i) \quad \Rightarrow \quad \mathbf{p}\dot{\mathbf{r}} = \sum_{i=1}^{3k} \left(\mathrm{m}\dot{r}_i^2\right) = 2T \ ,$$

so the Hamiltonian becomes

$$\mathrm{H} = \mathbf{p}\dot{\mathbf{r}} - \mathrm{L} = 2T - (T - U) = T + U \ .$$

Therefore, the Hamiltonian is given by the total energy of the system, as desired. □

Recalling Proposition 2.7, it is immediately evident that the equations of motion in a mechanical system in which energy is conserved can be obtained directly from the Hamiltonian, which is a first integral solution. Such systems are refered to as *Hamiltonian systems*, which are a primary topic in [Hai10].

Example 5.10 (The Two Body Problem)**.** Suppose two masses, M and m, are oriented such that the only force acting on them is that of mutual gravitational attraction. If we assume $\mathrm{M} >> \mathrm{m}$ and let the center of mass of the larger mass M be the origin of our coodinate system, then the the smaller mass m will follow an elliptical path in the plane containing both masses with the larger mass at one of its foci. Our goal is to determine the equation of motion of the smaller mass and show that its orbit is elliptical. Let $r : [0,t] \to \mathbb{R}$ and $\theta : [0,t] \to [0, 2\pi)$ in C^2 be the radial and angular positions of the smaller mass m in the plane, respectively. Then, the kinetic energy of the smaller mass is

$$T = \frac{1}{2}\mathrm{m}\left(\dot{r}^2 + r^2\dot{\theta}^2\right)$$

and its potential energy is

$$U = -\frac{\mathrm{GMm}}{r} \ ,$$

where G is the universal gravitational constant. Because there are no external forces acting on the system, the total energy is conserved. Thus, the Hamiltonian is a first integral solution. This gives us the the equation

$$\mathrm{H} = T + U = \frac{1}{2}\mathrm{m}\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - \frac{\mathrm{GMm}}{r} = C$$

$$\Rightarrow \quad \frac{(\mathrm{m}\dot{r})^2}{2\mathrm{m}} + \frac{\left(\mathrm{m}r^2\dot{\theta}\right)^2}{2\mathrm{m}r^2} - \frac{\mathrm{GMm}}{r} = C \ ,$$

for some constant C. Notice that this equation is equivalent to the equation

$$\frac{p_r^2}{2\mathrm{m}} + \frac{l_\theta^2}{2\mathrm{m}r^2} - \frac{\gamma}{r} = E \tag{23}$$

where p_r is the linear momentum, l_θ the angular momentum, E the total energy, and $\gamma \equiv \mathrm{GMm}$ the force constant. Noting that $\mathrm{L} = T - U$ is independent of θ, we have that θ is a cyclic coordinate, so l_θ is conserved. We obtain the values of the constant quantities E and l_θ from initial conditions, so we proceed by rewriting equation (23) in terms of the nonconstant quantites containing r and $\dot{r}$:

$$\frac{\mathrm{m}\dot{r}^2}{2} + \frac{l_\theta^2}{2\mathrm{m}r^2} - \frac{\gamma}{r} = E$$

$$\Rightarrow \quad \dot{r}^2 = \frac{2E}{\mathrm{m}} + \frac{2\gamma}{\mathrm{m}r} - \frac{l_\theta^2}{(\mathrm{m}r)^2} \quad .$$

We must manipulate $\dot{r}$ in order to continue. Notice that

$$\dot{r} = \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\dot{\theta} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\left(\frac{l_\theta}{\mathrm{m}r^2}\right) = \frac{l_\theta}{\mathrm{m}}\frac{\mathrm{d}\,(1/r)}{\mathrm{d}\theta} \quad ,$$

so we substitute and obtain

$$\left(\frac{l_\theta}{\mathrm{m}}\right)^2\left(\frac{\mathrm{d}\,(1/r)}{\mathrm{d}\theta}\right)^2 = \frac{2E}{\mathrm{m}} + \frac{2\gamma}{\mathrm{m}r} - \frac{l_\theta^2}{(\mathrm{m}r)^2}$$
$$\Rightarrow \quad l_\theta \int \frac{\mathrm{d}\,(1/r)}{\sqrt{2E\mathrm{m} + 2\gamma\mathrm{m}\,(1/r) - l_\theta^2\,(1/r)^2}} = \theta + \theta_0 \quad ,$$

where θ_0 is a constant of integration. Applying the change of variables $u = (1/r)$ yields

$$l_\theta \int \frac{\mathrm{d}u}{\sqrt{2E\mathrm{m} + 2\gamma\mathrm{m}u - l_\theta^2 u^2}} = \theta + \theta_0 \quad ,$$

which, after completing the square in the denominator of the integrand, has solution

$$u = \left(\frac{1}{l_\theta}\right)^2 \sqrt{2E\mathrm{m} + (\gamma\mathrm{m}/l_\theta)^2}\,\cos(\theta + \theta_0) - \gamma\mathrm{m} \quad .$$

Thus, we have that the position of the mass m is given by the equation

$$r = \frac{\alpha}{1 + \epsilon\cos(\theta + \theta_0)} \quad ,$$

where $\alpha \equiv (l_\theta)^2/\gamma\mathrm{m}$ and $\epsilon \equiv \sqrt{2E/\gamma^2\mathrm{m} + (1/l_\theta)^2}$. Without loss of generality, we can choose the intial condition $(r_0, \theta_0) = (\alpha, \pi/2)$, in which case the equation of motion can be written

$$r = \frac{\alpha}{1 + \epsilon\cos(\theta)} \quad . \tag{24}$$

Equation (24) describes a conic section with eccentricity ϵ. When the conic section is bounded, the mass m follows an elliptical orbit with mass M at one of its foci. Thus, the orbit is an ellipse whenever $0 \leq \epsilon < 1$. In particular, the orbit is circular when $\epsilon = 0$ and becomes more and more elongated as $\epsilon \to 1$. The unbounded case, whenever $\epsilon \geq 1$, describes a hyperbolic path which does not actually orbit the mass M. We conclude that the orbit of the smaller mass about the larger mass in our two body problem is indeed an ellipse with the larger mass at one of the foci. This concept, when applied to the more general two body problem, is known as Kepler's First Law.

The Lagrangian method can be applied to most classical mechanics problems and is often easier than working directly with Newton's equations of dynamics. Further discussion of its applications and more examples can be found in [Arn78].

6. Optimal Control

The theory of the calculus of variations has a very logical extension into the realm of control theory, in which we examine the behavior of dynamical systems subject to various constraints. If we concentrate on optimizing this behavior under a given set of constraints we can develop a set of rules for finding the solutions which yield optimal behavior, much like in the calculus of variations. This particular area of control theory is known as optimal control, pioneered by Russian mathematician Lev Pontryagin. Although we will not delve into the particulars of optimal control, we will look to understand its relationship to the calculus of variations and consider some of its applications. Proofs will be omitted, most of which can be found in [Lew06]. Before we begin, however, we must establish some notation which will motivate our discussion. Whenever possible, our notation here will mirror that used in our prior discussion of the calculus of variations.

Definition 6.1. Let $S \in \mathbb{R}^n$, $U \in \mathbb{R}^{n-1}$ be open sets and $f : S \times U \times \mathbb{R} \to \mathbb{R}^{n-1}$ be of class C^1. Then, the dynamical system represented by

$$\Sigma\left(S, f, U, t\right)$$

is called a *control system*. The sets S and U are called the *state space* and *control set*, respectively. Elements of the state space, $\mathbf{u} : \mathbb{R} \to \mathbb{R}^n$, are called *trajectories*, while those of the control set, $\mathbf{w} : \mathbb{R} \to \mathbb{R}^{n-1}$ are called *controls*. The pair $(\mathbf{u}, \mathbf{w})$ is called a *controlled trajectory*.

A control system is related to the differential equations

$$\dot{\mathbf{u}} = f\left(\mathbf{u}, \mathbf{w}, t\right) \quad , \tag{25}$$

and problems in optimal control involve solving equations (25) for an *optimal trajectory*, $(\mathbf{u}^*, \mathbf{w}^*)$, which yields optimal behavior in $\Sigma\left(S, f, U, t\right)$. Thus, the function f appears to play a role analogous to that of the time derivative $\dot{\mathbf{u}}$ in the calculus of variations. Naturally, we would like to define a Lagrangian for optimal control by simply substituting this function into the Lagrangian from variational calculus. We could then let the remainder of the theory follow directly from that of the calculus of variations. Although this approach illustrates the connection between the two theories, it does not provide the result we are looking for. However, we can still define some quantities which have similar meaning to their variational calculus counterparts.

Definition 6.2. Let $(\mathbf{u}, \mathbf{w})$ be a controlled trajectory for the control system $\Sigma\left(S, f, U, t\right)$. Then, the function $\mathrm{L}_\Sigma : S \times U \times \mathbb{R} \to \mathbb{R}$ of class C^1 is called the *control Lagrangian*, or *running cost*, for the system. The quantity

$$\mathrm{J}_\Sigma[(\mathbf{u}, \mathbf{w})] = T\left(\mathbf{u}, t_1\right) + \int_{t_0}^{t_1} \mathrm{L}_\Sigma\left(\mathbf{u}, \mathbf{w}, t\right)\ \mathrm{d}t$$

is called the *objective functional* for $\Sigma\left(S, f, U, t\right)$, where the function $T : S \times \mathbb{R} \to \mathbb{R}$ is the *terminal cost*.

By direct analogy to the calculus of variations, optimizing behavior in a control system requires optimizing the objective functional. However, we will do this over all admissible controlled trajectories rather than just all curves $\mathbf{u}$. As discussed in [Lew06], this inadvertently leads us toward analyzing the *control Hamiltonian*,

$$\mathrm{H}_\Sigma = \mathbf{p}\dot{\mathbf{u}} - \mathrm{L}_\Sigma \quad , \quad \mathbf{p} = \frac{\partial \mathrm{L}_\Sigma}{\partial \dot{\mathbf{u}}} \quad .$$

Unfortunately, this does not achieve the desired result, as a fundamental assumption of the calculus of variations is that the curve $\mathbf{u}$ be of class C^2. It turns out that this is too strong a restriction for curves in controlled trajectories. Instead, we must consider the *extended control Hamiltonian*,

$$\mathrm{H}_\Sigma^+ = \mathbf{h}\dot{\mathbf{u}} + \lambda \mathrm{L}_\Sigma \ \ ,$$

where $\mathbf{h} : \mathbb{R} \to \mathbb{R}^n$ is known as a *costate trajectory* and λ is a scalar. The costate trajectory is similar to a Lagrange multiplier, which we recall from the process of optimization of functions under given constraints in calculus. In particular, we are interested in the extended control Hamiltonian when the costate trajectory is an *adjoint response*. That is, when the costate trajectory satisfies the differential equations

$$\dot{\mathbf{u}} = \frac{\partial \mathrm{H}_\Sigma^+}{\partial \mathbf{h}}$$
$$\dot{\mathbf{h}} = -\frac{\partial \mathrm{H}_\Sigma^+}{\partial \mathbf{u}} \ \ .$$

We immediately notice that these equations resemble Hamilton's equations from the calculus of variations. The adjoint response and extended control Hamiltonian are essential to solving optimal control problems, but before we touch on this subject we must introduce one more new concept:

Definition 6.3. Let $\Sigma\left(S, f, U, t\right)$ be a control system and $\Phi : S \to \mathbb{R}^n$ be of class C^1 such that $S = \ker\left(\Phi\right)$ and $\mathrm{d}\Phi/\mathrm{d}t : S \to \mathbb{R}^{n-1}$ is a surjective map for all $\mathbf{u} \in S$. Then, S is called a *smooth control set*.

It is generally useful to study controlled trajectories where the initial and final states lie in the smooth control sets, $S_0 \subset S$ and $S_1 \subset S$, respectively.

We are now in a position to establish one of the major accomplishments of early control theory, Pontryagin's Maximum Principle. Much like the Necessary Condition of Euler-Lagrange from variational calculus, the Maximum Principle provides us with necessary conditions which must be satisfied by any optimal trajectory. Although these conditions are not sufficient, they are usually restrictive enough to allow intuition or numerical methods to determine which of the controlled trajectories found are actually optimal. We will merely state the theorem here; a proof can be found in [Lew06]. Our statement draws from those of both [Lew06] and [Nor07].

Theorem 6.4 (Pontryagin's Maximum Principle)**.** *Let $\Sigma\left(S, f, U, t\right)$ be a control system with control Lagrangian L_Σ, terminal cost T, and extended control Hamiltonian H_Σ^+ for $t \in [t_0, t_1]$. Futhermore, let $S_0 \subset S$ and $S_1 \subset S$ be smooth control sets such that $\mathbf{u}\left(t_0\right) \in S_0$ and $\mathbf{u}\left(t_1\right) \in S_1$ for any admissible controlled trajectory $\left(\mathbf{u}, \mathbf{w}\right)$. If $\left(\mathbf{u}^*, \mathbf{w}^*\right)$ is an optimal trajectory for $\Sigma\left(S, f, U, t\right)$, then there exists adjoint response $\mathbf{h}^*$ and scalar $\lambda \in \{0, -1\}$ such that the following conditions hold:*

(1) $\max_{\mathbf{w}\in U}\left[\mathrm{H}_\Sigma^+\left(\mathbf{u}^*, \mathbf{w}, \mathbf{h}^*, t\right)\right] \leq \mathrm{H}_\Sigma^+\left(\mathbf{u}^*, \mathbf{w}^*, \mathbf{h}^*, t\right)$ *for all* $t \in [t_0, t_1]$

(2) $\dot{\mathbf{h}}^* = -\left(\partial H_\Sigma^+/\partial \mathbf{u}\right)\big|_{\mathbf{u}=\mathbf{u}^*, \mathbf{w}=\mathbf{w}^*, \mathbf{h}=\mathbf{h}^*}$

(3) $\dot{\mathbf{u}}^* = \left(\partial H_\Sigma^+/\partial \mathbf{h}\right)\big|_{\mathbf{u}=\mathbf{u}^*, \mathbf{w}=\mathbf{w}^*, \mathbf{h}=\mathbf{h}^*}$

(4) $\left[\left(\mathbf{h} + \partial T/\partial \mathbf{u}\right)\mathbf{u}\right]\big|_{t_1} = 0$ *for all* $\mathbf{u} \in U$

(5) $\lambda^2 + \left\|\mathbf{h}^*\left(t_1\right)\right\|^2 \neq 0$

Moreover, whenever t is unconstrained, $\mathrm{H}_\Sigma^+(\mathbf{u}^*, \mathbf{w}^*, \mathbf{h}^*, t) = 0$ *for all t.*

Conditions (1) through (3) can also be expressed as

(1) $0 = \partial \mathrm{H}_\Sigma^+ / \partial \mathbf{w}$
(2) $\dot{\mathbf{h}} = -\partial \mathrm{H}_\Sigma^+ / \partial \mathbf{u}$
(3) $\dot{\mathbf{u}} = \partial \mathrm{H}_\Sigma^+ / \partial \mathbf{h}$

which clearly shows how each of the variables are related. However, the reformulation of (1) requires that the extended control Hamiltonian achieves its maximum within the interval (t_0, t_1), which is not always the case. Conditions (4) and (5) are the boundary conditions. Notice that the terminal cost T plays the role of the differentiable map Φ in the definition of a smooth control set. Thus, condition (4) states that the adjoint response $\mathbf{h}$ is orthogonal to the state space S at the final state of the system. This relationship is also true at the intial state, since S_0 is a smooth control set. Condition (5) states that the adjoint response at the final state and the scalar λ cannot both be zero. This condition is sometimes refered to as the nontriviality condition for the total adjoint vector, $\lambda^2 + \|\mathbf{h}^*(t_1)\|^2$.[4]

Pontryagin's Maximum Principle is one of the triumphs of mathematics over the last half century. Although we cannot fully appreciate its utility here, it is evident that it provides a powerful method for solving optimal control problems. The Maximum Principle can be applied in a vast array of subjects which utilize optimal control, anything from economics and financial mathematics to mechanics and natural dynamical systems. Several examples of these applications can be found in [Nor07].

Example 6.5 (Optimal Mining)**.** Suppose a miner purchases the rights to a mine for a set period of time and wishes to make the largest profit possible within that time period. The miner sells the mined ore at a constant price, but the instantaneous value of the ore decreases at the rate at which it is extracted from the mine. The miner also incurs a cost equal to the squared rate of extraction divided by the amount of ore remaining in the mine. Our goal is to determine the extraction rate which will maximize the miner's profits. Let τ be the amount of time the miner owns the rights to the mine, p the constant price at which the ore is sold, u the amount of ore in the mine, and w the extraction rate. Then, the miner's cost can be expressed by a control system $\Sigma(S, f, U, t)$ with $u \in S$, $w \in U$, $f = -w$, and $t \in [0, \tau]$. Suppose $u(0) = X$ is the initial amount of ore in the mine and that the miner places no value on any ore left in the mine at time τ. Thus, we know that the control Lagrangian is given by

$$\mathrm{L}_\Sigma = \frac{w^2}{u} - pw$$

and the terminal cost is $T = 0$. This means that to maximize the miner's profit we must minimize the objective functional,

$$\mathrm{J}_\Sigma[u, w] = \int_0^\tau \left(\frac{w^2}{u} - pw \right) \mathrm{d}t \quad .$$

To do so, we will use Pontryagin's Maximum Principle. Consider the extended control Hamiltonian,

$$\mathrm{H}_\Sigma^+ = h\dot{u} + \lambda \mathrm{L}_\Sigma = -hw + \lambda \left(\frac{w^2}{u} - pw \right) \quad ,$$

[4]Here, $\|\cdot\|$ denotes the standard Euclidean norm.

where h is the adjoint response. Condition (4) of Theorem 6.4 states that $h(\tau)\,u(\tau) = 0$ for all u, which necessarily requires that $h(\tau) = 0$ since $u(\tau)$ is not necessarily zero. Thus, condition (5) requires that $\lambda = -1$. Applying conditions (1) through (3) with $\lambda = -1$ then yields the equations

$$\begin{aligned} 0 &= \frac{\partial \mathrm{H}_\Sigma^+}{\partial w} = -h + p - \frac{2w}{u} \\ \dot{h} &= -\frac{\partial \mathrm{H}_\Sigma^+}{\partial u} = -\frac{w^2}{u^2} \\ \dot{u} &= \frac{\partial \mathrm{H}_\Sigma^+}{\partial h} = -w \ . \end{aligned}$$

The first two equations give us that

$$\frac{w}{u} = \frac{p-h}{2} \quad \Rightarrow \quad \dot{h} = -\frac{1}{4}(p-h)^2 \ ,$$

which is a separable differential equation that we can solve to obtain the adjoint response. Writing $\dot{h}$ as $\mathrm{d}h/\mathrm{d}t$, we have

$$\begin{aligned} &\int (p-h)^2 \ \mathrm{d}h = -\frac{1}{4}\int \mathrm{d}t \\ \Rightarrow \quad & h^3 - 3ph^2 + 3p^2h = -\frac{3}{4}t + C \ , \end{aligned}$$

where C is a constant of integration. The condition $h(\tau) = 0$ allows us to determine that $C = 3\tau/4$, so we have

$$h^3 - 3ph^2 + 3p^2h - \frac{3}{4}(t-\tau) = 0 \ .$$

This equation has one real solution for h:

$$h = p - \left(p^3 - \frac{3}{4}(t-\tau)\right)^{1/3} \ .$$

Returning to the first condition, we see that the control is given by

$$w = \frac{u(p-h)}{2} \ .$$

Thus, the optimal trajectory (u^*, w^*) must satisfy the equation

$$w^* = \frac{1}{2}\,u^*\left(p^3 - \frac{3}{4}(t-\tau)\right)^{1/3} \ . \tag{26}$$

Therefore, the extraction rate which will maximize the miner's profit can be found by analyzing controlled trajectories which satisfy equation (26) with initial condition $u(0) = X$.

We hinted earlier that the methods of optimal control require weaker assumptions on the domain of the objective functional than does the calculus of variations. We will now conclude our discussion of the Maximum Principle by illustrating how its relatively relaxed differentiability requirements, compared to those of the Euler-Lagrange equations, give it an advantage over the classical techniques of the calculus of variations when determining candidate curves for optimization.

Example 6.6 (The Brachistochrone and Optimal Control)**.** Suppose we reformulate the setup of Example 2.6 as an optimal control problem. Let the endpoints of the wire again be $(0,0)$, (x_1, y_1), where $x_1 > 0$, $y_1 > 0$, and our coordinate system defines the downward direction as $+y$. Now, however, let $\mathbf{u} : [0, t_1] \to S \subset \mathbb{R}^2$ be of class C^1 satisfying the boundary conditions $\mathbf{u}(0) = (0,0)$ and $\mathbf{u}(t_1) = (x_1, y_1)$. If we denote $\mathbf{u} = (u_1, u_2)$, then we have the equation

$$\dot{u}_1^2 + \dot{u}_2^2 = 2\mathrm{g}u_2 \ ,$$

which we have adapted from the time Lagrangian derived in Example 2.6. If we define the control $\mathbf{w} = (w_1, w_2)$ by the relations $\dot{u}_1 = w_1\sqrt{2\mathrm{g}u_2}$, $\dot{u}_2 = w_2\sqrt{2\mathrm{g}u_2}$ for $\mathbf{w} \in \{(w_1, w_2)\,|w_1^2 + w_2^2 = 1\}$, which we will call the set U, then it is easy to verify that $\mathbf{w}$ satisfies this time Lagrangian equation. Let $\Sigma(S, f, U, t)$ be a control system for the controlled trajectories $(\mathbf{u}, \mathbf{w})$, where

$$f = \left(h_1 w_1 \sqrt{2\mathrm{g}u_2}, h_2 w_2 \sqrt{2\mathrm{g}u_2}\right)$$

for adjoint response $\mathbf{h} = (h_1, h_2)$. Notice that the control Lagrangian is given by $\mathrm{L}_\Sigma = 1$ and the terminal cost is $T = 0$. Thus, in order to determine the shape of the wire which minimizes the time it takes for the bead to travel its length, we must to minimize the objective functional

$$\mathrm{J}_\Sigma[\mathbf{u}, \mathbf{w}] = \int_0^{t_1} \mathrm{d}t \ .$$

To do so, we will use Pontryagin's Maximum Principle. Consider the extended control Hamiltonian,

$$\mathrm{H}_\Sigma^+ = \mathbf{h}\dot{\mathbf{u}} + \lambda \mathrm{L}_\Sigma = (h_1 w_1 + h_2 w_2)\sqrt{2\mathrm{g}u_2} + \lambda \ .$$

Condition (4) of Theorem 6.4 states that the relations

$$\begin{aligned} h_1(t_1)\, u_1(t_1) &= 0 \\ h_2(t_1)\, u_2(t_1) &= 0 \end{aligned}$$

are satisfied for all u_1 and u_2, which necessarily requires that $h_1(t_1) = h_2(t_1) = 0$ since $u_1(t_1)$, $u_2(t_1)$ are nonzero by assumption. Thus, condition (5) requires that $\lambda = -1$. Applying conditions (2) and (3) with $\lambda = -1$ then yields the equations

$$\begin{aligned} \dot{h}_1 &= -\frac{\partial \mathrm{H}_\Sigma^+}{\partial u_1} = 0 \\ \dot{h}_2 &= -\frac{\partial \mathrm{H}_\Sigma^+}{\partial u_2} = -\frac{\mathrm{g}(h_1 w_1 + h_2 w_2)}{\sqrt{2\mathrm{g}u_2}} \\ \dot{u}_1 &= \frac{\partial \mathrm{H}_\Sigma^+}{\partial h_1} = w_1\sqrt{2\mathrm{g}u_2} \\ \dot{u}_2 &= \frac{\partial \mathrm{H}_\Sigma^+}{\partial h_2} = w_2\sqrt{2\mathrm{g}u_2} \ . \end{aligned}$$

Condition (1) requires that the extended control Hamiltonian be maximized over the control set, which occurs when

$$\begin{aligned} w_1 &= \frac{h_1}{\sqrt{h_1^2 + h_2^2}} \\ w_2 &= \frac{h_2}{\sqrt{h_1^2 + h_2^2}} \ , \end{aligned}$$

although this is not immediately obvious (it can be verified by noting that $w_1^2+w_2^2=1$ is still valid). Thus, we have that

$$\dot{h}_2 = -\frac{\mathrm{g}\,(h_1w_1+h_2w_2)}{\sqrt{2\mathrm{g}u_2}} = -\frac{\mathrm{g}\sqrt{h_1^2+h_2^2}}{\sqrt{2\mathrm{g}u_2}} \ .$$

Since $\dot{u}_1$ is nonzero by assumption, we can use a series of substitutions to write $u_2' = \mathrm{d}u_2/\mathrm{d}u_1$ as

$$\begin{aligned} u_2' &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1} = \frac{\dot{u}_2}{\dot{u}_1} = \frac{w_2}{w_1} = \frac{h_2}{h_1} \\ &\Rightarrow \quad 1+(u_2')^2 = \frac{h_1^2+h_2^2}{h_1^2} \ , \end{aligned}$$

and because h_1 is constant we have that

$$u_2'' = \frac{1}{h_1}\frac{\mathrm{d}h_2}{\mathrm{d}u_1} = \frac{\dot{h}_2}{h_1\dot{u}_1} = -\frac{\sqrt{h_1^2+h_2^2}}{2u_2h_1w_1} = -\frac{h_1^2+h_2^2}{2u_2h_1^2} = \frac{-1}{2u_2}\left(\frac{h_1^2+h_2^2}{h_1^2}\right) \ .$$

It is now evident that $2u_2u_2'' = -\left(1+(u_2')^2\right)$, so we obtain

$$2u_2u_2'' + (u_2')^2 + 1 = 0 \ .$$

This is exactly equation (9) from Example 2.6, whose solution is a cycloid. However, this time we did not require that u_2 be of class C^2. It turns out there are solutions of class C^1 which we left out in Example 2.6, called *spurious solutions*. These solutions, which cannot be obtained via the methods of the calculus of variations, are indeed viable minimizing curves, illustrating the advantages of optimal control and the Maximum Principle over their variational calculus counterparts. For a discussion of the spurious solutions to the brachistochrone problem, see [SW02].

References

[Arn78] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer-Verlag, New York, 1978, Translated from the Russian by K. Vogtmann and A. Weinstein, Graduate Texts in Mathematics, 60. MR 0690288 (57 #14033b)

[Ben08] Martin Bendersky, *The calculus of variations*, December 2008, Lecture notes.

[Bol61] Oskar Bolza, *Lectures on the calculus of variations*, 2nd ed. Chelsea Publishing Co., New York, 1961. MR 0123206 (23 #A534b)

[GF63] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Revised English edition translated and edited by Richard A. Silverman, Prentice-Hall Inc., Englewood Cliffs, N.J., 1963. MR 0160139 (28 #3353)

[Hai10] Ernst Hairer, *Hamiltonian systems*, January-February 2010, Lecture notes.

[Lew06] Andrew D. Lewis, *The maximum principle of pontryagin in control and in optimal control*, May 2006, Lecture notes.

[Nor07] J.R. Norris, *Optimal control*, November 2007, Lecture notes.

[Olv12] Peter J. Olver, *The calculus of variations*, January 2012, Lecture notes.

[Ric96] V. Frederick Rickey, *History of the brachistochrone*, May 1996, Lecture notes.

[SW02] Héctor J. Sussmann and Jan C. Willems, *The brachistochrone problem and modern control theory*, Contemporary trends in nonlinear geometric control theory and its applications (México City, 2000), World Sci. Publ., River Edge, NJ, 2002, pp. 113–166. MR 1881487 (2002k:93024)