The Calculus of Variations

Advisor: Christina Tonnesen-Friedman

Author: Erin Whitney

Fall 2011 - Winter 2012

Chapter 1

An Introduction

1.1 Motivation

The Calculus of Variations is a branch of mathematics that has commonly been used as a tool for other areas of study, including but not limited to physics, geometry, and economics. This area of mathematics involves finding the extrema of functionals:

Definition 1.1.1. *A functional is a mapping from a set of functions, S, to the real numbers, $\mathbb{R}$.*

Calculus of Variations is used as a device for optimization problems that involve functionals, as opposed to elementary calculus involving functions. It is thus the case that we are working with extrema that are functions instead of vectors. These functionals often involve definite integrals.

One application of this branch of mathematics is one in differential geometry; finding the geodesic, or the shortest curve, i.e. the curve with minimum arclength, connecting any two given points. The most intuitive answer, and sometimes the correct answer, is a straight line between the two points. However, if the two points lie on a curved surface, a straight line might not be possible. It follows that we need another device to find the geodesics of this particular class of curves where a straight line is not the solution.

The set up to find geodesics requires that we parametrize the surface with which we are working. We let the surface of interest be defined by Σ, and let it be described by the position vector function $r : \sigma \to \mathbb{R}^3$. Here, σ is a nonempty, connected, open subset of $\mathbb{R}^2$ and $(u, v) \in \sigma$. We use the following parametrization and assume that x, y, and z are smooth functions of

$(u,v) \in \sigma$:
$$r(u,v) = (x(u,v), y(u,v), z(u,v)),$$
where r is a 1-1 and onto function of σ onto Σ. We refine our attention to smooth, simple curves, γ, on Σ. It follows that there is a curve $\tilde{\gamma}$ in σ that, under r, maps to γ. Assume that $\tilde{\gamma}$ is parametrized by $(u(t), v(t))$ where $t_0 \leq t \leq t_1$. So, for the purpose of this motivating example, we will define the components of the first fundamental form and leave the details up to the reader, which can be found in reference [1]. We define the components as follows:
$$E = \left|\frac{\partial r}{\partial u}\right|^2, \; F = \frac{\partial r}{\partial u} \cdot \frac{\partial r}{\partial v}, \; G = \left|\frac{\partial r}{\partial v}\right|^2.$$
The arclength is thus defined by:
$$L(\gamma) = \int_{t_0}^{t_1} \sqrt{Eu'^2 + 2Fu'v' + Gv'^2} dt$$
where the functions u and v are functions of t. We will now demonstrate this abstract concept with the following concrete example:

Example 1.1.1. *We will now consider an example involving finding the geodesics on a unit sphere. First of all, we can parametrize an octant of a unit sphere, Σ. This parameterization is as follows:*
$$r(u,v) = (\sin u \cos v, \sin u \sin v, \cos u)$$
such that
$$\sigma = \{(u,v) \,|\, 0 < u < \frac{\pi}{2}, \, 0 < v < \frac{\pi}{2}\}.$$
We can then see that the components of the first fundamental form, E, F, and G, are as follows:
$$E = |(\cos u \cos v, \cos u \sin v, -\sin u)|^2 = 1$$
$$F = (\cos u cosv, cosusinv, -\sin u) \cdot (-\sin u \sin v, \sin u \cos v, 0) = 0$$
$$G = |(-\sin u \sin v, \sin u \cos v, 0)|^2 = \sin^2 u.$$
It then follows that the arclength is defined as
$$L(\gamma) = \int_{t_0}^{t_1} \sqrt{u'^2 + sin^2 uv'^2} dt.$$
We can use calculus of variations to determine the minimum of this arclength function as γ spans the set of all curves from $\gamma(t_0) = \gamma_0$ to $\gamma(t_1) =$

γ_1, that is, $(u(t), v(t))$ spans the set of all functions $(u, v) : [t_0, t_1] \to \sigma$ s.t. $(u(t_0), v(t_0)) = (u_0, v_0)$ and $(u(t_1), v(t_1)) = (u_1, v_1)$. The mathematics covered in this paper will make doing so possible [6].

In the remainder of the text, we will examine specific applications and problems that utilize the calculus of variations.

1.2 Background Information

Let X be a set of functions, a function space, defined on the real numbers, $\mathbb{R}$. Next, we will define the infinite vector space that we are working within by use of the following definitions, which all come out of reference [6]:

Definition 1.2.1. *A normed vector space is a vector space in which a norm is defined. The norm is a real-valued function on the vector space whose value at f, denoted $\|f\|$, satisfies the following properties:*
(1) $\|f\| \geq 0$
(2) $\|f\| = 0 \iff f = 0$
(3) $\|\alpha f\| = |\alpha| \|f\|$
(4) $\|f + g\| \leq \|f\| + \|g\|$.

Definition 1.2.2. *An inner product on a vector space Y is a function on $Y \times Y$, denoted $\langle \cdot, \cdot \rangle$, such that the following statements hold $\forall\ f, g, h \in Y$, $\forall\ \alpha \in \mathbb{R}$:*
(1) $\langle f, f \rangle \geq 0$
(2) $\langle f, f \rangle = 0 \iff f = 0$
(3) $\langle f + g, h \rangle = \langle f, g \rangle + \langle g, h \rangle$
(4) $\langle f, g \rangle = \langle g, f \rangle$
(5) $\langle \alpha f, g \rangle = \alpha \langle f, g \rangle$.

A vector space with an inner product structure leads us to the definition of a norm on that vector space, and this norm is defined as:

$$||f|| = \sqrt{\langle f, f \rangle}.$$

Definition 1.2.3. *A Banach space is a complete normed vector space, i.e., a normed vector space in which every Cauchy sequence in the space converges.*

Definition 1.2.4. *A Hilbert Space is a special type of Banach space in that it has an inner product structure.*

For the purpose of this paper, we will be working within Hilbert Spaces. Note that the function space denoted by $C^{(n)}[x_0,\, x_1]$ is the set of functions with at least an n^{th} order continuous derivative on the interval $[x_0,\, x_1]$. This is indeed an infinite dimensional vector space. To view it as a subset of a Hilbert Space, one could introduce the following inner product:

$$\int_{x_0}^{x_1} f(x)g(x)dx = 0.$$

The following theorems, from reference [6], offer a brief review of topics covered in calculus and are necessary to arriving at the remainder of our results.

Theorem 1.2.1. (Mean Value Theorem) *Let $x_0, x_1 \in \mathbb{R}$ be such that $x_0 < x_1$. Let f be a function continuous in $[x_0,\, x_1]$ and differentiable in $(x_0,\, x_1)$. Then $\exists\ \xi$ such that $x_0 < \xi < x_1$ and $f(x_1) = f(x_0) + (x_1 - x_0)f'(\xi)$.*

Theorem 1.2.2. (Taylor's Theorem) *Let f be a function with the first n derivatives continuous in $[x_0, x_1]$ such that $f^{(n+1)}(x)$ exists $\forall\ x \in (x_0, x_1)$. Then, $\exists$ a number $\xi \in (x_0, x_1)$ such that:*

$$\begin{aligned} f(x_1) =& f(x_0) + (x_1 - x_0)f'(x_0)\frac{(x_1 - x_0)^2}{2}f''(x_0) + \dots \\ &\dots + \frac{(x_1 - x_0)^n}{n!}f^{(n)}(x_0) + \frac{(x_1 - x_0)^{n+1}}{(n+1)!}f^{(n+1)}(\xi) \end{aligned}$$

We define the set S as follows:

$$S = \{y \in C^{(2n)}[x_0, x_1] \,|\, y^{(k)}(x_0) = y_0^k \, and \, y^{(k)}(x_1) = y_1^k \, for \, k = 0, \dots, n-1\}$$

such that y_0^k and y_1^k are fixed constants. For functions $\hat{y} \in S$ in an ϵ - neighborhood of a function $y \in S$, that is, $||\hat{y} - y|| < \epsilon$, $\exists\ \eta \in X$ s.t.

$$\hat{y} = y + \epsilon\eta.$$

We define the set H by:

$$H = \{\eta \in C^{(n)}[x_0, x_1] \,|\, y + \epsilon\eta \in S\}.$$

We will examine functionals of the form

$$J(y) = \int_{x_0}^{x_1} f(x, y, y', \dots, y^{(n)})dx.$$

Definition 1.2.5. *Let J: $X \to \mathbb{R}$ be a functional defined on the function space $(X, \|\cdot\|)$. Let $S \subset X$. The functional J is said to have a local maximum in S at $y \in S$ if $\exists\, \epsilon > 0$ s.t. $J(\hat{y}) - J(y) \leq 0\ \forall\, y \in S$ s.t. $\|\hat{y} - y\| < \epsilon$. The functional is said to have a local minimum in S at $y \in S$ if y is a local maximum in S for $-J$.*

Definition 1.2.6. *The functional J is said to have a local extremum, extrema if plural, in S at $y \in S$ if it is either a local maximum or a local minimum.*

Remark: In this text, we will first consider the case in which $n = 1$, and then we will consider the case in which $n = 2$.

1.3 First Case: n=1

Theorem 1.3.1. *Let J: $C^2[x_0, x_1] \to \mathbb{R}$ be a functional of the form:*

$$J(y) = \int_{x_0}^{x_1} f(x, y, y')dx,$$

where f has continuous partial derivatives of second order with respect to x, y, y' and $x_0 < x_1$. If $y \in S$ is an extremum for J, then

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0 \tag{1.1}$$

$\forall\, x \in [x_0, x_1]$.

Proof. The proof of Theorem 1.3.1 can be found in reference [2]. □

Definition 1.3.1. *Equation* (1.1) *is called the Euler-Lagrange Equation.*

Definition 1.3.2. *If $y \in S$ satisfies the Euler-Lagrange Equation for J then we say J is stationary at the point y, that is, y is an extremal for J. In this case, y is not necessarily a local extremum of J.*

In the next section we will prove a more general case with $n = 2$.

1.4 Second Case: n=2

Theorem 1.4.1. *Consider functionals of the form:*

$$J(y) = \int_{x_0}^{x_1} f(x, y, y', y'')dx$$

such that f has continuous partial derivatives of third order with respect to x, y, y',y'', $y \in C^4[x_0, x_1]$, and with endpoint conditions such that $y(x_0) = y_0$, $y'(x_0) = y'_0$, $y(x_1) = y_1$, and $y'(x_1) = y'_1$. We have the following result:

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) = 0. \tag{1.2}$$

Proof. In this case, we have that the set S is defined by:

$$S = \{y \in C^4[x_0, x_1] | y(x_0) = y_0, y'(x_0) = y'_0, y(x_1) = y_1, y'(x_1) = y'_1\}$$

and the set H is defined by:

$$H = \{\eta \in C^4[x_0, x_1] | \eta(x_0) = \eta'(x_0) = \eta(x_1) = \eta'(x_1) = 0\}.$$

Now, suppose J has a local extremum in S at $y \in S$, and let $\hat{y} = y + \epsilon\eta$. We then consider $J(\hat{y}) - J(y)$. Taylor's Theorem tells us that:

$$\begin{aligned} f(x, \hat{y}, \hat{y}', \hat{y}'') &= f(x, y + \epsilon\eta, y' + \epsilon\eta', y'' + \epsilon\eta'') \\ &= f(x, y, y', y'') + \epsilon\left(\eta\frac{\partial f}{\partial y} + \eta'\frac{\partial f}{\partial y'} + \eta''\frac{\partial f}{\partial y''}\right) + O(\epsilon^2). \end{aligned}$$

It follows that:

$$\begin{aligned} J(\hat{y}) - J(y) &= \int_{x_0}^{x_1} f(x, \hat{y}, \hat{y}', \hat{y}'')dx - \int_{x_0}^{x_1} f(x, y, y', y'')dx \\ &= \int_{x_0}^{x_1} \left(f(x, y, y', y'') + \epsilon\left\{\eta\frac{\partial f}{\partial y} + \eta'\frac{\partial f}{\partial y'} + \eta''\frac{\partial f}{\partial y''}\right\} + O(\epsilon^2)\right) \\ &\quad - f(x, y, y', y'')dx \\ &= \epsilon\int_{x_0}^{x_1} \left(\eta\frac{\partial f}{\partial y} + \eta'\frac{\partial f}{\partial y'} + \eta''\frac{\partial f}{\partial y''}\right) dx + O(\epsilon^2) \\ &= \epsilon\delta J(\eta, y) + O(\epsilon^2). \end{aligned}$$

The first variation for this function is thus

$$\delta J(\eta, y) = \int_{x_0}^{x_1} \left(\eta\frac{\partial f}{\partial y} + \eta'\frac{\partial f}{\partial y'} + \eta''\frac{\partial f}{\partial y''}\right) dx.$$

If J has a local extremal at $y \in S$ then we get that

$$\delta J(\eta, y) = \int_{x_0}^{x_1} \left(\eta \frac{\partial f}{\partial y} + \eta' \frac{\partial f}{\partial y'} + \eta'' \frac{\partial f}{\partial y''} \right) dx = 0 \tag{1.3}$$

$\forall\ \eta \in H$. Now, Integration by parts yields:

$$\begin{aligned}\int_{x_0}^{x_1} \eta'' \frac{\partial f}{\partial y''} dx &= \eta' \frac{\partial f}{\partial y''}\bigg|_{x_0}^{x_1} - \int_{x_0}^{x_1} \eta' \frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right) dx \\ &= -\eta \frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right)\bigg|_{x_0}^{x_1} + \int_{x_0}^{x_1} \eta \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) dx \\ &= \int_{x_0}^{x_1} \eta \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) dx.\end{aligned}$$

Note that we have used the boundary conditions that were given above. Using integration by parts again gives us:

$$\begin{aligned}\int_{x_0}^{x_1} \eta' \frac{\partial f}{\partial y'} dx &= \eta \frac{\partial f}{\partial y'}\bigg|_{x_0}^{x_1} - \int_{x_0}^{x_1} \eta \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx \\ &= -\int_{x_0}^{x_1} \eta \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx\end{aligned}$$

Once again, by using the given boundary conditions, we get that equation (1.3) gives us that:

$$\int_{x_0}^{x_1} \eta \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) \right] dx = 0,$$

$\forall\ \eta \in H$. Now, by assumption that $y \in C^4[x_0, x_1]$, we see that f has continuous third order partial derivatives so that the expression:

$$E(x) = \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right)$$

must be continuous $\forall\ y \in C^4[x_0, x_1]$ on the interval $[x_0, x_1]$. By Lemma 1.4.1, which can be found below, we see that y must satisfy:

$$\frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{\partial f}{\partial y} = 0,$$

the Euler-Lagrange equation. It follows that $y \in S$ is an extremum for the functional J, and we are done. □

Lemma 1.4.1. *Assume that* $\langle \eta, g \rangle = 0 \ \forall \ \eta \in H$. *If* $g : [x_0, x_1] \to \mathbb{R}$ *is a continuous function then* $g = 0$ *on the interval* $[x_0, x_1]$.

Proof. The proof follows from a suitable adjustment of the proof of Lemma 2.2.2, which can be found in reference [6]. □

We have laid the foundation of this branch of mathematics. In the next chapter, we take these foundational tools and use them to consider a special case utilizing some basic knowledge of convex functions.

Chapter 2

Convex Sets and Functions

2.1 Motivation: Special Case

Consider the set of smooth functions K such that
(i) $y(x) > 0$, $-1 < x < 1$
(ii) $y(\pm 1) = 0$
(iii) $y'(-1) = \frac{2}{w_S}$, $y'(1) = \frac{-2}{w_N}$
where $1 \leq w_S \leq w_N \in \mathbb{Z}$ such that w_S and w_N are relatively prime, i.e., the greatest common divisor of w_S and w_N is 1. We want to find a function $y \in K$ such that it minimizes the following integral:

$$\int_{-1}^{1} (y(x)'')^2 dx.$$

Thus, the functional that we will be working with is the map

$$y(x) \to \int_{-1}^{1} (y(x)'')^2 dx, \tag{2.1}$$

and it is inspired by Eugenio Calabi, as explained in reference [2].

Theorem 2.1.1. *If $y \in K$ minimizes functional* (2.1) *then*

$$y'' = Az + B$$

for some constants A and B.

Proof. We want to use the Euler-Lagrange equation to determine for which functions it is satisfied, i.e., to determine which functions might minimize functional (2.1). The Euler-Lagrange equation (1.2) reduces to:

$$\frac{d^2}{dx^2}(2y'') = 0.$$

Thus, since the second derivative with respect to x of $2y''$ is 0, we have that y'' is a linear function of x. It follows that $y'' = Ax + B$ for some constants A and B, as desired. □

Now we examine a corollary that follows from the last theorem.

Corollary 2.1.1. *If $y \in K$ minimizes functional* (2.1) *then*

$$y(x) = (1 - x^2)\left[\frac{\left(\frac{1}{w_N} - \frac{1}{w_S}\right)z + \left(\frac{1}{w_S} + \frac{1}{w_N}\right)}{2}\right]. \tag{2.2}$$

Proof. The proof of this corollary is a direct computational result from Theorem 2.1.1. Since we have that $y'' = Ax + B$ for some constants A and B and that $y(\pm 1) = 0$, we get that $y(x) = (1 - x^2)(cx + d)$ for constants c and d. But, since $y'(-1) = \frac{2}{w_S}$ and $y'(1) = \frac{-2}{w_N}$, we get the desired result, that is, we get function (2.2). □

Now if $w_S = w_N = 1$ then we can easily show that this is indeed a minimizer by applying the Schwarz Inequality.

Theorem 2.1.2. (The Schwarz Inequality) *[6] Let f_1 and f_2 be any two real-valued functions in [a,b] that are integrable. The Schwarz Inequality is given by the following:*

$$\left[\int_a^b f_1(x)f_2(x)dx\right]^2 \leq \int_a^b [f_1(x)]^2 dx \int_a^b [f_2(x)]^2 dx.$$

So, we apply the Schwarz Inequality to functional (2.1) for $w_S = w_N = 1$ in order to show that $y = 1 - x^2$ minimizes it. Let $f_1(x) = 1$ and $f_2(x) = y''$ for $y \in K$. We compute:

$$\int_{-1}^{1} 1^2 dx \int_{-1}^{1} (y'')^2 dx \geq \left[\int_{-1}^{1} y'' dx\right]^2$$

This implies that:

$$\begin{aligned}\int_{-1}^{1} (y'')^2 dx &\geq \frac{\left[\int_{-1}^{1} y'' dx\right]^2}{\int_{-1}^{1} 1^2 dx} \\ &= \frac{([y']_{-1}^{1})^2}{1 - (-1)} \\ &= \frac{(-2-2)^2}{2} \\ &= 8\end{aligned}$$

Now, we compute the left hand side of the inequality when $y = 1 - x^2$:

$$\begin{aligned}\int_{-1}^{1} (y'')^2 dx &= \int_{-1}^{1} (-2)^2 dx \\ &= [4x]_{-1}^{1} \\ &= [4 - (-4)] \\ &= 8\end{aligned}$$

Therefore, we have shown that the function $y = 1 - x^2$ does minimize the functional.

2.2 General Case

For w_S and w_N not both equal to 1, the Schwarz Inequality is sharp, i.e. equality does not hold, for function (2.2). Thus, we cannot use Theorem 2.1.2 to conclude that function (2.2) is a minimizer, and therefore, other tools must be applied. These next two definitions are the building blocks for the tools that we will need to do this.

Definition 2.2.1. *A set $\Omega \subset \mathbb{R}^3$ is convex if the line segments connecting any two points, z_1 and z_2, are elements of Ω, i.e., if z_1, $z_2 \in \Omega$ then it follows that*

$$w(t) = (1 - t)z_1 + tz_2 \in \Omega$$

$\forall\ t \in [0/1]$.

Definition 2.2.2. *Let $\Omega \subset \mathbb{R}^2$ be a convex set. A function $f\colon \Omega \to \mathbb{R}$ is said to be convex if*

$$f(w(t)) = f((1 - t)z_1 + tz_2) \leq (1 - t)f(z_1) + tf(z_2) \tag{2.3}$$

$\forall\ z_1, z_2 \in \Omega$ *and* $\forall\ t \in [0, 1]$.

The Mean Value Theorem implies that $\exists$ a $\tau \in$ (0,1) such that

$$f(w(t)) = f(z_1) + t(z_2 - z_1) \cdot \nabla f(w(\tau)).$$

Then, this equation along with the definition of a convex function implies that

$$(z_2 - z_1) \cdot \nabla f(z_1) \leq f(z_2) - f(z_1)$$

$\forall\ t \in$ (0,1). Since $0 < \tau < t$, the partial derivatives of f are continuous, and τ depends on t. If z_1 is a stationary point for f, we get that $\nabla f(z_1) = 0$ and thus the inequality reduces to

$$f(z_1) \leq f(z_2)$$

$\forall\ z_2 \in \Omega$. Hence, for convex functions, a stationary point is a minimum for f in Ω.

From this, we can explore a sufficient condition for a functional to have a minimum at an extremal. We have the following theorem:

Theorem 2.2.1. *Let the set Ω_x be convex for each $x \in [x_0,x_1]$, and let f be a convex function of the variables $(y,y',y'') \in \Omega_x$. If y is a smooth extremal for J then J has a minimum at y for the fixed endpoint problem.*

Proof. Now, we consider

$$J(y) = \int_{x_0}^{x_1} f(x, y, y', y'')dx.$$

Here, we let D_f denote the domain of the function f and let the set

$$\Omega_x = \{(y, y', y'') \in \mathbb{R}^3 | (x, y, y', y'') \in D_f\}$$

$\forall\ x \in [x_0,\ x_1]$ be convex. Also, let f be a convex function on the set Ω_x. By inequality (2.3), it follows that:

$$\begin{aligned} f(x, \hat{y}, \hat{y}', \hat{y}'') - f(x, y, y', y'') \geq \quad & (\hat{y} - y) f_y(x, y, y', y'') \\ & + (\hat{y}' - y') f_{y'}(x, y, y', y'') \\ & + (\hat{y}'' - y'') f_{y''}(x, y, y', y'') \end{aligned}$$

for any points (y,y',y''), $(\hat{y},\ \hat{y}',\ \hat{y}'') \in \Omega_x$. Now we consider the end point parameters $\hat{y}(x_0) = y(x_0)$ and $\hat{y}(x_1) = y(x_1)$, i.e., y is a smooth extremal for J and $\hat{y} \in S$. We see that:

$$\begin{aligned} J(\hat{y}) - J(y) &= \int_{x_0}^{x_1} (f(x, \hat{y}, \hat{y}', \hat{y}'') - f(x, y, y', y''))dx \\ &\geq \int_{x_0}^{x_1} (\hat{y} - y) f_y(x, y, y', y'') + (\hat{y}' - y') f_{y'}(x, y, y', y'') \\ &\qquad + (\hat{y}'' - y'') f_{y''}(x, y, y', y'')dx. \end{aligned}$$

Using integration by parts yields the two following equations:

$$\int_{x_0}^{x_1} (\hat{y}' - y') f_{y'}(x, y, y', y'') dx = -\int_{x_0}^{x_1} \frac{d}{dx} f_{y'}(x, y, y', y'')(\hat{y} - y) dx.$$

$$\int_{x_0}^{x_1} (\hat{y}'' - y'') f_{y''}(x, y, y', y'') dx = -\int_{x_0}^{x_1} \frac{d}{dx} f_{y''}(x, y, y', y'')(\hat{y}' - y') dx.$$

We now have that:

$$J(\hat{y}) - J(y) \geq \int_{x_0}^{x_1} (\hat{y} - y) f_y - \frac{d}{dx}(f_{y'})(\hat{y} - y) - \frac{d}{dx}(f_{y''})(\hat{y}' - y') dx.$$

Using integration by parts once again gives us:

$$J(\hat{y}) - J(y) \geq \int_{x_0}^{x_1} (\hat{y} - y) f_y - \frac{d}{dx}(f_{y'})(\hat{y} - y) - \frac{d^2}{dx^2}(f_{y''})(\hat{y} - y) dx.$$

It follows that:

$$J(\hat{y}) - J(y) \geq \int_{x_0}^{x_1} (\hat{y} - y) \left(f_y - \frac{d}{dx}(f_{y'}) - \frac{d^2}{dx^2}(f_{y''}) \right) dx$$

and since y satisfies the Euler-Lagrange Equation, we get that:

$$J(\hat{y}) - J(y) \geq 0.$$

This implies that J has a minimum at y, as desired Thus, we are done. □

We now return back to the example. Let $\Omega_x = \mathbb{R}^3$, which is certainly convex. We also have that $f(y, y', y'') = (y'')^2$ is convex since it is the graph of a parabola, which is clearly convex. Thus, by Theorem 2.2.1, we can see that $y(x) = (1 - x^2)(cx + d)$ is indeed a minimizer.

Chapter 3

The Hamiltonian Formulation

3.1 Background Information

In Chapter 1, we delved into the general case for the Euler-Lagrange equation, used as a tool in the calculus of variations. Now, we will examine a system of $2n$ first-order ordinary differential equations, known as Hamilton's equations, that yield equivalent results but while utilizing a system that comes along with less complications. The following two definitions set us up for this new process.

Definition 3.1.1. *A point transformation from one point $(x, y(x))$ to another point $(X, Y(X))$ is composed of relations of the following forms:*

$$X = X(x, y),$$

and

$$Y = Y(x, y).$$

Definition 3.1.2. *A contact transformation is a transformation such that the functions are dependent on the derivative of the dependent variable.*

We will now examine a specific type of transformation. Let $y : [x_0, x_1] \to \mathbb{R}$ be a smooth function such that $y''(x) \neq 0$. Define p as follows:

$$p = y'(x). \tag{3.1}$$

Since we assumed that $y''(x) \neq 0$, that is,

$$\frac{dp}{dx} \neq 0,$$

we may solve for x in terms of p in equation (3.1). This fact follows directly from the Inverse Function Theorem, which is stated below:

Theorem 3.1.1. The Inverse Function Theorem *[6] Let $f : U \to \mathbb{R}^n$ be a smooth map defined on an open subset U of $\mathbb{R}^n$ $(n \geq 1)$. Assume that, at some point $x_0 \in U$, the Jacobian matrix $J(f)$ is invertible. Then, there is an open subset V of $\mathbb{R}^n$ and a smooth map $g : V \to \mathbb{R}^n$ such that*
(i) $y_0 = f(x_0) \in V$
(ii) $g(y_0) = x_0$
(iii) $g(V) \subset U$
(iv) $g(V)$ is an open subset of $\mathbb{R}^n$
(v) $f(g(y)) = y \ \forall \ y \in V$
In particular, $g : V \to g(V)$ and $f : g(V) \to V$ are inverse bijections.

We then define a new function H such that:

$$H(p) = -y(x) + px, \tag{3.2}$$

where x is viewed as a function of p. We see that equation (3.1) and equation (3.2) give us a transformation from $(x, y(x))$ to $(p, H(p))$. This is an example of this particular type of transformation, known as the Legendre transformation. One of the qualities that this type of transformation possesses that makes it important is that it is an involution, which is defined below.

Definition 3.1.3. *An involution is a transformation that is its own inverse.*

Now, we will go through the steps to confirm that this is in fact the case with this Legendre transformation. So, we look at equation (3.2) and take the derivative with respect to p :

$$\begin{aligned}\frac{dH}{dp} &= \frac{d}{dp}\left(-y(x)\right) + \frac{d}{dp}\left(px\right) \\ &= -\frac{dy}{dx}\frac{dx}{dp} + p\frac{dx}{dp} + x \\ &= (-y'(x) + p)\frac{dx}{dp} + x \\ &= x \ (\textit{by Equation } (3.1))\end{aligned}$$

We also see that:

$$\begin{aligned}-H(p) + xp &= -(-y(x) + px) + xp \\ &= y(x).\end{aligned}$$

Thus, we have verified that the Legendre transformation is in fact an example of an involution. In other words, when we apply the transformation to the pair $(p, H(p))$, we yield the pair $(x, y(x))$, which is the original pair.

As we have in the past, we will now look at functions of the form $f(x, y, y')$. We define the variable p in a new way:

$$p = \frac{\partial f}{\partial y'} \tag{3.3}$$

where we have assumed that

$$\frac{\partial^2 f}{\partial y'^2} \neq 0.$$

So, similarly to before, we may solve equation (3.3) for y' in terms of p (and also x and y).

Definition 3.1.4. *Here, y' is the only variable actively involved in the transformation, and is appropriately known as an active variable. The other two variables are known as the passive variables.*

We redefine H in this case as follows:

$$H(x, y, p) = -f(x, y, y') + py'. \tag{3.4}$$

As before, we will verify that this is an involution. We take the derivative of Equation (3.4) with respect to p.

$$\begin{aligned}
\frac{dH}{dp} &= \frac{d}{dp}(-f(x, y, y')) + \frac{d}{dp}(y'p) \\
&= -\frac{df}{dy'}\frac{dy'}{dp} + p\frac{dy'}{dp} + y' \\
&= \left(-\frac{df}{dy'} + p\right)\frac{dy'}{dp} + y' \\
&= y' \ (by\, Equation\, (3.3)).
\end{aligned}$$

We also have that:

$$\begin{aligned}
-H(x, y, p) + y'p &= -(-f(x, y, y') + py') + y'p \\
&= f(x, y, y').
\end{aligned}$$

Thus, we have, once again, verified that this is indeed an involution. Now, we will take a look at a few examples in order to see more tangible applications.

Example 3.1.1. *Let* $f(x, y, y') = \sqrt{1 + y'^2}$*. Using equation* (3.3)*, we get that*

$$p = \frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2}}.$$

Solving for y' *yields:*

$$y' = \frac{p}{\sqrt{1 - p^2}}.$$

Thus, $H(x, y, p)$ *is as follows:*

$$\begin{aligned} H(x, y, p) &= -\sqrt{1 + y'^2} + y'p \\ &= -\frac{1}{\sqrt{1 - p^2}} + \frac{p^2}{\sqrt{1 - p^2}} \\ &= -\sqrt{1 - p^2}. \end{aligned}$$

We now see that

$$\frac{dH}{dp} = \frac{p}{\sqrt{1 - p^2}} = y',$$

and that

$$-H(x, y, p) + py' = \sqrt{1 + y'^2} - y'p + py' = f(x, y, y'),$$

as desired.

We will now look at a different example.

Example 3.1.2. *We want to find the Hamiltonian* H *for the following function:*

$$f(x, y, y') = \sqrt{\frac{1 + y'^2}{y}}.$$

So, we get that p *is as follows:*

$$p = \frac{\partial f}{\partial y'} = \frac{y'}{y\sqrt{\frac{1+y'^2}{y}}}.$$

It follows that:

$$y' = \sqrt{\frac{p^2 y}{1 - p^2 y}}.$$

So, we get that the Hamiltonian H is:

$$\begin{aligned} H(x,y,p) &= -f(x,y,y') + y'p \\ &= -\sqrt{\frac{1+y'^2}{y}} + y'p \\ &= -\sqrt{\frac{y}{1-p^2y}} + p^2\sqrt{\frac{y}{1-p^2y}} \\ &= (p^2-1)\sqrt{\frac{y}{1-p^2y}}. \end{aligned}$$

In the next section, we will examine a new Legendre transformation where we expand our area of interest to the n-dimensional case, and apply it to the case where f is the integrand of some functional.

3.2 Hamilton's Equations

We are now interested in a smooth function, $L(t, q, \dot{q})$, s.t. $\mathbf{q} = (q_1, q_2, \ldots, q_n)$. We define p to be:

$$p_k = \frac{\partial L}{\partial \dot{q_k}}, \tag{3.5}$$

where $k = 1, 2, \ldots, n$, and the variables $\dot{q_k}$ are functions of p, q, and t. Using the Implicit Function Theorem, we can show that equation (3.5) can be solved for $\dot{q}$ in terms of p (and t and q), as long as the Hessian matrix, as follows:

$$M_L = \begin{pmatrix} \frac{\partial^2 L}{\partial \dot{q_1}\partial \dot{q_1}} & \frac{\partial^2 L}{\partial \dot{q_1}\partial \dot{q_2}} & \cdots & \frac{\partial^2 L}{\partial \dot{q_1}\partial \dot{q_n}} \\ \frac{\partial^2 L}{\partial \dot{q_2}\partial \dot{q_1}} & \frac{\partial^2 L}{\partial \dot{q_2}\partial \dot{q_2}} & \cdots & \frac{\partial^2 L}{\partial \dot{q_2}\partial \dot{q_n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 L}{\partial \dot{q_n}\partial \dot{q_1}} & \frac{\partial^2 L}{\partial \dot{q_n}\partial \dot{q_2}} & \cdots & \frac{\partial^2 L}{\partial \dot{q_n}\partial \dot{q_n}} \end{pmatrix}$$

is nonsingular. In other words, we assume the condition:

$$det M_L \neq 0.$$

For the n-dimensional case, we define H as follows:

$$H(t,q,p) = -L(t,q,\dot{q}) + \sum_{k=1}^{n} \dot{q_k} p_k \tag{3.6}$$

Definition 3.2.1. *The function H is known as the Hamiltonian Function and the function L is known as the Lagrangrian.*

Once again, we check to see that we truly have an involution. So, we take the derivative of equation (3.6) with respect to p_k to get:

$$\begin{aligned}\frac{\partial H}{\partial p_k} &= \sum_{j=1}^{n}\left(-\frac{\partial L}{\partial \dot{q_j}} + p_j\right)\frac{\partial \dot{q_j}}{\partial p_k} + \dot{q_k} \\ &= \dot{q_k}.\end{aligned}$$

We also get that:

$$-H(t,q,p) + \sum_{k=1}^{n} \dot{q_k} p_k = L(t,q,\dot{q}).$$

Thus, we have verified that we do indeed have an involution. We will now look at an example:

Consider the functional J of the form

$$J(q) = \int_{t_0}^{t_1} L(t,q,\dot{q})dt,$$

such that $q = (q_1, q_2, \ldots, q_n)$ and L is a smooth function satisfying the following:

$$det M_L \neq 0.$$

Note that we let k take on the values $1, 2, \ldots, n$. In the case that q is a smooth extremal for J, we have the Euler-Lagrange Equations:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q_k}} - \frac{\partial L}{\partial q_k} = 0. \tag{3.7}$$

This is a generalization of section (1.1) to the case of several dependent variables. We can apply the Legendre Transformation to Equation (3.7) to get:

$$p_k = \frac{\partial L}{\partial \dot{q_k}}, \tag{3.8}$$

and

$$\dot{q_k} = \frac{\partial H}{\partial p_k}. \tag{3.9}$$

Since t and q are passive variables, we have that

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}, \tag{3.10}$$

and

$$\frac{\partial H}{\partial q_k} = -\frac{\partial L}{\partial q_k}.$$

Since we know that q is an extremal, we know that it satisfies the Euler-Lagrange Equation, i.e., it satisfies Equation (3.7). Thus, we have that

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q_k}} = \frac{\partial L}{\partial q_k} = \dot{p_k}.$$

From Equation (3.10), we get that

$$\dot{p_k} = -\frac{\partial H}{\partial q_k}. \tag{3.11}$$

The fact that the Legendre Transformation is an involution implies that solving the system of n Euler-Lagrange Equations is the same as solving the system of $2n$ equations, i.e., Equation (3.9) and Equation (3.11).

For the following example, we will assume that we are in the special case where $n = 1$, that is, we will assume that we are working within $\mathbb{R}$ as opposed to $\mathbb{R}^n$. This will help clarify the example and avoid the problem of getting lost with the notation.

Example 3.2.1. *The Lagrangian for a linear harmonic oscillator is*

$$L(t, q, \dot{q}) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2,$$

where m is mass and k is the restoring force coefficient [6]. Let us compute p and then find $H(t, q, p)$. Using equation (3.8)*, we get that*

$$p = m\dot{q}.$$

It follows that

$$\dot{q} = \frac{p}{m}.$$

By equation (3.6)*, we then see that:*

$$\begin{aligned} H(t, q, p) &= \frac{-1}{2}m\dot{q}^2 + \frac{1}{2}kq^2 + \dot{q}p \\ &= \frac{-1}{2m}p^2 + \frac{1}{2}kq^2 + \frac{p^2}{m} \\ &= \frac{1}{2m}p^2 + \frac{1}{2}kq^2. \end{aligned}$$

Secondly, we find the Euler-Lagrange equation. Using equation (3.7)*, we get*

$$\frac{d}{dt}m\dot{q} + kq = 0.$$

Thus,

$$m\ddot{q} + kq = 0.$$

Then, we find Hamilton's equations:

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{1}{m}p$$

and

$$\dot{p} = \frac{-\partial H}{\partial q} = -kq. \tag{3.12}$$

It follows that

$$\ddot{q} = \frac{1}{m}\dot{p}.$$

Thus,

$$m\ddot{q} = \dot{p}. \tag{3.13}$$

Equation (3.12) *combined with Equation* (3.13) *gives us that*

$$m\ddot{q} = -kq,$$

which is exactly the Euler-Lagrange equation. Thus, we have shown that the two methods yield equivalent results.

Now, we will look at a very different example involving a more tangible, physical topic.

Example 3.2.2. *Let* $(x(z), y(z), z)$*, where* $z \in [z_0, z_1]$*, represent a space curve, denoted by* γ*. The optical path length of* γ *for a medium with a refractive index of* $n(x, y, z)$ *is as follows:*

$$J(x, y) = \int_{x_0}^{x_1} n(x, y, z)\sqrt{1 + x'^2 + y'^2}dz.$$

In order for J *to be a light ray, by Fermat's Principal,* γ *must be an extremum. In order to keep the notation consistent, we will let* $q_1 = x$*,* $q_2 = y$*,* $z = t$*,* $z_0 = t_0$*, and* $z_1 = t_1$*. We also define* L *to be*

$$L(t, q, \dot{q}) = n(t, q)\sqrt{1 + |\dot{q}|^2}.$$

This is known as the optical Lagrangian. We note that $q = (q_1, q_2)$ *and* $n(t, q) = n(x, y, z)$*. We then get that*

$$p_k = \frac{\partial L}{\partial \dot{q_k}} = \frac{n\dot{q}}{\sqrt{1 + |\dot{q_k}|^2}},$$

which represents what is known as the generalized momenta. We now have that:

$$p_1^2 + p_2^2 - n^2 = -\frac{n^2}{1 + |\dot{q}|^2}.$$

It thus follows that

$$\begin{aligned} \dot{q_k} &= \frac{p_k}{n}\sqrt{1 + |\dot{q_k}|^2} \\ &= \frac{p_k}{n}\sqrt{\frac{n^2}{n^2 - p_1^2 - p_2^2}} \\ &= \frac{p_k}{n^2 - p_1^2 - p_2^2}. \end{aligned}$$

Then, the Hamiltonian function is as follows:

$$\begin{aligned} H(t, q, p) &= \sum_{k=1}^{2} p_k \dot{q_k} - L(t, q, \dot{(q)}) \\ &= \frac{p_1^2 + p_2^2}{\sqrt{n^2 - p_1^2 - p_2^2}} - \frac{n^2}{\sqrt{n^2 - p_1^2 - p_2^2}} \\ &= -\sqrt{n^2 - p_1^2 - p_2^2}. \end{aligned}$$

Finally, we can determine Hamilton's equations for this example. They are as follows:

$$\dot{q_k} = \frac{\partial H}{\partial p_k} = \frac{p_k}{\sqrt{n^2 - p_1^2 - p_2^2}}$$

and

$$\dot{p_k} = -\frac{\partial H}{\partial q_k} = \frac{1}{\sqrt{n^2 - p_1^2 - p_2^2}}\frac{\partial n}{\partial q_k}.$$

Since additional tools are needed to solve this system, this example will be continued, and thus completed, in a later section [6].

So, we now have seen that we have two methods, using the Euler-Lagrange equations and using the Hamiltonian function H, to arrive at the same result. It is important to recognize exactly why all this was done, that is, to recognize the motivation behind the math. Some complications that come along with the Euler-Lagrange Equations are that they are second-order and frequently have non-linear first derivatives. The new functions are first order, and the system is able to be derived from one generating function, known as the Hamiltonian. The following two definitions introduce this method in a formal manner.

Definition 3.2.2. *The Hamiltonian system is the system of* $2n$ *equations,* (3.9) *and* (3.11) *in the text above.*

Definition 3.2.3. *The equations that this system is comprised of are known as the Hamiltonian equations.*

Therefore, using this new method comes along with many added benefits and functions that are easier to work with.

3.3 Symplectic Maps

In this section, we will examine a new transformation. We will assume that we are in the case such that $n = 1$. From the last section, we have that H is a function of the variables t, q, and p such that:

$$H(t, q, p) = -L(t, q, \dot{q}) + \dot{q}p,$$

where $L(t, q, \dot{q})$ is a smooth function and

$$p = \frac{\partial L}{\partial \dot{q}}.$$

We also have that:

$$\dot{q} = \frac{\partial H}{\partial p},$$

and

$$\dot{p} = -\frac{\partial H}{\partial q}.$$

Now, we will consider a smooth function, Φ, of variables t, q, and Q such that:

$$\frac{\partial^2 \Phi}{\partial q \partial Q} \neq 0,$$

and define $\hat{H}$ by:

$$\hat{H}(t, Q, P) = H(t, q, p) + \frac{\partial \Phi}{\partial t}. \tag{3.14}$$

where P and Q are related to p and q as follows:

$$p = \frac{\partial \Phi}{\partial q}, \tag{3.15}$$

and

$$P = -\frac{\partial \Phi}{\partial Q}. \tag{3.16}$$

Note that since we are making the assumption that

$$\frac{\partial^2 \Phi}{\partial q \partial Q} \neq 0,$$

it follows that

$$\frac{d}{dq}\left(\frac{\partial \Phi}{\partial Q}\right) = -\frac{\partial P}{\partial q} \neq 0, \tag{3.17}$$

and

$$\frac{d}{dQ}\left(\frac{\partial \Phi}{\partial q}\right) = \frac{\partial p}{\partial Q} \neq 0. \tag{3.18}$$

We can use these facts, along with the Implicit Function Theorem, to verify that we have a transformation. From Equation (3.17), we may solve Equation (3.16) for q in terms of P and Q. Then, from Equation (3.15) we have p in terms of P and Q as well. Likewise, from Equation (3.18), we can solve for Q in terms of p and q and then Equation (3.16) gives us P in terms of p and q. Thus, we have that $(t, p, q) \to (t, P, Q)$ is a transformation.

Taking the derivative of the function Φ with respect to t gives us:

$$\frac{d}{dt}\Phi(t, q, Q) = \frac{\partial \Phi}{\partial q}\dot{q} + \frac{\partial \Phi}{\partial Q}\dot{Q} + \frac{\partial \Phi}{\partial t}.$$

If we substitute in for $\frac{\partial \Phi}{\partial t}$, using Equation (3.14), then we see that

$$\frac{d\Phi}{dt} = \frac{\partial \Phi}{\partial q}\dot{q} + \frac{\partial \Phi}{\partial Q}\dot{Q} + \hat{H}(t, Q, P) - H(t, q, p).$$

Then, by equation (3.15) and (3.16), we see that:

$$\frac{d\Phi}{dt} = p\dot{q} - P\dot{Q} + \hat{H}(t, Q, P) - H(t, q, p). \tag{3.19}$$

It follows directly from Equation (3.19) that:

$$p\dot{q} - H(t,q,p) = P\dot{Q} - \hat{H}(t,Q,P) + \frac{d}{dt}\Phi(t,Q,P). \tag{3.20}$$

Let us now define two functionals J and $\hat{J}$ by:

$$J(q,p) = \int_{t_0}^{t_1} p\dot{q} - H(t,q,p)dt$$
$$\hat{J}(Q,P) = \int_{t_0}^{t_1} P\dot{Q} - \hat{H}(t,Q,P)dt.$$

Since Φ is a perfect differential, and the integrands of the two functionals J and $\hat{J}$ only differ by this derivative, J and $\hat{J}$ are variationally equivalent.

Definition 3.3.1. *Two functionals are said to be variationally equivalent if they yield the same set of extrema.*

Due to Equation (3.20), this is the case here. Thus, we have that the Euler-Lagrange equations must be equivalent. However, the Euler-Lagrange equations are similarly the Hamiltonian equations:

$$\dot{q} = \frac{\partial H}{\partial p} \tag{3.21}$$

$$\dot{p} = -\frac{\partial H}{\partial q}, \tag{3.22}$$

and

$$\dot{Q} = \frac{\partial \hat{H}}{\partial P} \tag{3.23}$$

$$\dot{P} = -\frac{\partial \hat{H}}{\partial Q}. \tag{3.24}$$

Therefore, the function $(q,p) \to (Q,P)$ transforms Equation (3.22) and Equation (3.21) into Equation (3.23) and Equation (3.24). Thus, we have a symplectic map, defined as follows.

Definition 3.3.2. *[6] A symplectic map is a transformation of the form:*

$$Q = Q(t,q,p)$$
$$P = P(t,q,p),$$

such that the transformation takes the Hamiltonian system, Equation (3.22) *and Equation* (3.21)*, and transforms it into another Hamiltonian system, as follows:*

$$\dot{Q} = \frac{\partial \hat{H}}{\partial P},$$
$$\dot{P} = -\frac{\partial \hat{H}}{\partial Q}.$$

Now, we will see this concept on a more concrete level, through use of an example.

Example 3.3.1. *In this example, we will take a look at the harmonic oscillator from Example* (3.2.1)*. In this case, the Hamiltonian is, as we found out before, the following:*

$$H = \frac{p^2}{2m} + \frac{1}{2}kq^2, \tag{3.25}$$

where q represents the position of a particle of mass m at time t, p represents the momentum, and ω represents a constant relating to the restoring force. Also recall that the Hamiltonian system for this is:

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m}$$
$$\dot{p} = -\frac{\partial H}{\partial q} = -kq.$$

We choose Φ to be:

$$\Phi(q, Q) = \frac{\sqrt{km}q^2}{2} cotQ.$$

Without explanation, this Φ seems to be chosen at random. Since this is not the case, we will take a look at the reasoning and process behind making this choice. Let $f(P)$ be a function that depends entirely on P. We want $\hat{H}$ to depend only on P, so we are looking to satisfy

$$\hat{H}(P, Q) = \frac{(f(P))^2}{2m}.$$

[3]. We are also looking for p and q such that

$$p = f(P)cosQ$$
$$q = \frac{f(P)}{\sqrt{km}} sinQ.$$

It follows that

$$\frac{p}{q} = \sqrt{km}cotQ.$$

Thus

$$p = q\sqrt{km}cotQ.$$

Since $p = \frac{\partial\Phi}{\partial q}$, *integrating with respect to* q *yields the* Φ *that we defined above. We will now continue with the details of the example.*

Now, we can compute p *and* P *using a different method, yielding equivalent results, as follows:*

$$p = \frac{\partial\Phi}{\partial q} = q\sqrt{km}cotQ. \tag{3.26}$$

and

$$P = -\frac{\partial\Phi}{\partial Q} = \frac{\sqrt{km}q^2}{2sin^2Q}. \tag{3.27}$$

From Equation (3.27), *we have*

$$q = \sqrt{\frac{2P}{\sqrt{km}}}sinQ,$$

and from Equation (3.26) *and Equation* (3.27), *we see that*

$$p = \sqrt{2P\sqrt{km}}cosQ.$$

Thus, we get that

$$f(P) = \sqrt{2P\sqrt{km}}.$$

We can then find the new transformation, which is in fact transformed from the Hamiltonian. It is as follows:

$$\hat{H}(t, Q, P) = \sqrt{\frac{k}{m}}P.$$

It follows that the Hamiltonian system is:

$$\dot{Q} = \frac{\partial\hat{H}}{\partial P} = \sqrt{\frac{k}{m}}$$

$$\dot{P} = -\frac{\partial\hat{H}}{\partial Q} = 0.$$

Now, we can solve this system. We get that

$$Q = \sqrt{\frac{k}{m}}t + c_1$$
$$P = c_2,$$

where c_1 and c_2 are constants. This yields

$$q = \sqrt{\frac{2c_2}{\sqrt{km}}} sin\left(\frac{k}{m}t + c_1\right),$$
$$p = \sqrt{2c_2\sqrt{km}} cos\left(\frac{k}{m}t + c_1\right).$$

Therefore, we have solved the system, and we are done [6].

In the next section, we will explore a Hamiltonian system that is easier to work with.

3.4 The Hamiltonian-Jacobi Equation

The previous section works as a precursor for this next concept. Say we are interested in producing a generating function $\Phi(t, q, Q)$ such that the transformed Hamiltonian is zero, that is, $\hat{H}(t, Q, P) = 0$. It follows that we get the Hamiltonian system below:

$$\dot{Q} = \frac{\partial \hat{H}}{\partial P} = 0,$$
$$\dot{P} = -\frac{\partial \hat{H}}{\partial Q} = 0.$$

It is clear that from this yields:

$$Q = \alpha$$
$$P = \beta$$

where α and β are constants. Also, from Equation (3.14) along with the assumption that $\hat{H} = 0$ we get that

$$H(t, q, p) + \frac{\partial \Phi}{\partial t} = 0.$$

We now have an equation of three variables. We have that H is a function of t, q, and p, so we work to eliminate the variable p. In order to do this, we use Equation (3.15) to get:

$$H\left(t, q, \frac{\partial \Phi}{\partial q}\right) + \frac{\partial \Phi}{\partial t} = 0. \tag{3.28}$$

This equation is a first-order partial differential equation for the generating function Φ.

Definition 3.4.1. *This first-order partial differential equation is called the <u>Hamiltonian-Jacobi equation</u>.*

There are a few characteristics of the Hamiltonian-Jacobi equation that are important to note. First of all, Φ is not directly present in the equation. The equation only depends on the partial derivatives of Φ. However, none of these partial derivatives are with respect to Q. The reason for this is that the Q only provide us with "initial data" [6].

It follows from our discussion in Section 3.3 that in order for any solution of the Hamilton-Jacobi equation to be a true generating function for a symplectic transformation, we also need to assume that it is complete, defined below.

Definition 3.4.2. *A solution $\Phi(t, q, \alpha)$ is <u>complete</u> if it has continuous second derivatives and*

$$\frac{\partial^2 \Phi}{\partial q \partial \alpha} \neq 0.$$

The following theorem summarizes what we have discussed in this section thus far.

Theorem 3.4.1. *[6] Suppose that $\Phi(t, q, \alpha)$ is a complete solution to the Hamilton-Jacobi equation (3.28). Then, the general solution to the Hamiltonian system*

$$\dot{q} = \frac{\partial H}{\partial p}$$
$$\dot{p} = -\frac{\partial H}{\partial q}$$

is given by the equations

$$\frac{\partial \Phi}{\partial \alpha} = -\beta, \tag{3.29}$$

and

$$\frac{\partial \Phi}{\partial q} = p, \tag{3.30}$$

where β is an arbitrary constant.

Proof. See pages 177 through 179 in Reference [6] for the complete proof of this theorem. □

Instead of including the proof in its entirety, we will discuss the result and how it came to be in a less rigid way. So, we first make the assumption that $\Phi(t, q, \alpha)$ is a complete solution to the Hamilton-Jacobi equation, labeled Equation (3.28). Since Φ generates a symplectic transformation in which $\hat{H} = 0$, we have that Equation (3.23) and Equation (3.24) are solved by $Q = \alpha$ and $P = \beta$. Then Equation (3.29) and Equation (3.30) follow directly from Equation (3.15) and Equation (3.16).

Reference [6] provides the reader with a sequential list of steps in order to get the solution based on the Hamilton-Jacobi equation:

(1) Find the Hamiltonian H.
(2) Determine the Hamilton-Jacobi equation.
(3) Find the complete solution Φ to the Hamilton-Jacobi equation.
(4) Form Equation (3.29) for β, a constant.
(5) Solve these equations for q to obtain the general solution.

We will now take a look at an example. Note that in this example we are in the case that $n = 2$. However, everything for this case follows naturally from the previous discussion.

Example 3.4.1. *This example uses information from Example (3.2.2). The Hamiltonian for the path of a light ray from this previous example is*

$$H(t, q, p) = -\sqrt{n^2 - p_1^2 - p_2^2}.$$

From Equation (3.28)*, we have that the Hamilton-Jacobi equation is*

$$-\sqrt{n^2 - \left(\frac{\partial \Phi}{\partial q_1}\right)^2 - \left(\frac{\partial \Phi}{\partial q_2}\right)^2} + \frac{\partial \Phi}{\partial t} = 0.$$

By subtracting by $\frac{\partial \Phi}{\partial t}$*, squaring, and then adding* $\left(\frac{\partial \Phi}{\partial q_1}\right)^2 + \left(\frac{\partial \Phi}{\partial q_2}\right)$ *to each side, we get the following equation:*

$$\left(\frac{\partial \Phi}{\partial q_1}\right)^2 + \left(\frac{\partial \Phi}{\partial q_2}\right)^2 + \left(\frac{\partial \Phi}{\partial t}\right)^2 = n^2.$$

We can also write this equation using the following notation:

$$|\nabla \Phi|^2 = n^2,$$

where ∇ *is the symbol for the gradient. We now suppose that* $n = \mu\sqrt{q_1}$*, where* n *denotes the refractive index and* $\mu > 1$ *is a constant. Additionally, we assume that* $q_1 \geq 1$*. Substituting the correct information for* n *yields:*

$$\left(\frac{\partial \Phi}{\partial q_1}\right)^2 + \left(\frac{\partial \Phi}{\partial q_2}\right)^2 + \left(\frac{\partial \Phi}{\partial t}\right)^2 = \mu^2 q_1.$$

We then can determine what Φ *is for this example. We get the following solution for the previous equation:*

$$\Phi(t, q, \alpha) = \frac{2}{3\mu^2}(\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2))^{3/2} + \alpha_1 q_2 + \alpha_2 t.$$

Then, we see that the matrix $M := \left(\frac{\partial^2 \Phi}{\partial q_i \partial \alpha_j}\right)$ *is:*

$$M = \begin{pmatrix} \frac{-\alpha_1}{\sqrt{\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2)}} & \frac{-\alpha_2}{\sqrt{\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2)}} \\ 1 & 0 \end{pmatrix}$$

Thus, we can compute the determinant of the square matrix as follows:

$$detM = -\frac{-\alpha_2}{\sqrt{\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2)}}.$$

Generalizing Definition (3.4.2) to the case in which $n = 2$*, we see that this is complete, that is, the determinant is nonzero, when*

$$\alpha_1^2 + \alpha_2^2 < \mu^2 q_1$$

and $\alpha_2 \neq 0$*. Now, we get*

$$\beta_1 = -\frac{\partial \Phi}{\partial \alpha_1} = \frac{\alpha\sqrt{\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2)}}{\mu^2} - q_2$$
$$\beta_2 = -\frac{\partial \Phi}{\partial \alpha_2} = \frac{\alpha_2\sqrt{\mu^2 q_1 - (\alpha_1^2 + \alpha_2^2)}}{\mu^2} - t.$$

Now, we are able to solve for q_1 and q_2. We get

$$q_1(t, \alpha, \beta) = \frac{\mu^2}{\alpha_2^2}(\beta_2 + t)^2 + \frac{\alpha_1^2 + \alpha_2^2}{\mu^2}$$
$$q_2(t, \alpha, \beta) = \frac{\alpha_1}{\alpha_2}t + \beta_2 - \beta_1.$$

Thus, we are done [6].

In this next section, we will explore a different application of this field of mathematics.

3.5 Further Exploration in Differential Geometry

Another field besides theoretical mathematics in which symplectic structures appear in is physics. We will describe the motion of a system by the curve $\gamma = P(t)$ where P is a point on an n-dimensional real manifold Q.

Definition 3.5.1. *Here, Q is called the configuration space and P is described by the local coordinates $q_1, q_2, \ldots q_n$, which are known as the position variables.*

The basic principle behind physical differential geometry problems has to do with the principle of least action. We are looking to find the curve of least distance. So, we assume that the system has a Lagrangian function such that

$$L = L(q, \dot{q}, t).$$

In the physical sense, this function is defined as

$$L = E_{kin} - E_{pot}$$

where E_{kin} denotes the kinetic energy of the system and E_{pot} denotes the potential energy of the system. By the principal of least action, we want to find the curve γ that minimizes

$$\int_{t_0}^{t_1} L(q, \dot{q}, t)dt = \int_{t_0}^{t_1} (E_{kin} - E_{pot})dt.$$

Since we have stated that γ is the minimum curve, it must satisfy the Euler-Lagrange equation:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0$$

where $i = 1, \ldots, n$. Now, we can, as discussed before, take a look at Hamilton's equations for this special case. We transform from coordinates representing position and velocity, $(q, \dot{q})$, to coordinates representing position and momentum, (q, p). As described in section (3.2), the transformation is

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

where $i = 1, \ldots, n$. This is a Legendre Transformation between tangent and cotangent bundles, defined below.

Definition 3.5.2. *The tangent bundle is the disjoint union of the tangent spaces of a differential manifold. The cotangent bundle is the disjoint union of the dual of the tangent spaces of a differential manifold.*

This transformation takes us from the space TQ to the phase space T^*Q. In this case, the Hamiltonian function H can be defined as

$$\begin{aligned} H = H(p, q, t) &= p\dot{q} - L \\ &= m\dot{q}^2 - E_{kin} + E_{pot} \\ &= E_{kin} + E_{pot} \end{aligned}$$

since $E_{kin} = \frac{m\dot{q}^2}{2}$ and $p = m\dot{q}$. We can also use the Hamilton-Jacobi equations in this case, just as discussed in section (3.4).

We will now take a symplectic look at this case. We can look to H to define a Hamiltonian vector space X_H on the phase space T^*Q. This vector space is defined as follows:

$$X_H := \sum_i \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \sum \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i}.$$

We now consider the possibility of integral curves $\gamma(t) = (q(t), p(t))$.

Definition 3.5.3. *[1] An integral curve, γ, is a curve in which its tangent vectors, denoted $\dot{\gamma}$, are equal to the given vector of the vector field of that point at every point of the curve. This is often written as*

$$\dot{\gamma}(t) = X_H(\gamma(t)).$$

Since we have that

$$\dot{\gamma} = \dot{q}_i \frac{\partial}{\partial q_i} + \dot{p}_i \frac{\partial}{\partial p_i},$$

we can then see that $\gamma(t)$ being an integral curve of X_H is equivalent to Hamilton's Equations:

$$\frac{\partial H}{\partial p_i} = \dot{q}_i$$
$$\frac{\partial H}{\partial q_i} = -\dot{p}_i.$$

We can now look at this with respect to differential forms. We have that

$$dH = \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i.$$

There exists a 2 form such that

$$\omega = \sum_i dq_i \wedge dp_i \in \Omega^2.$$

This is a so-called symplectic form. Then

$$\omega(\dot{\gamma}(t)) = \dot{q}dp - \dot{p}dq.$$

We can then view Hamilton's equations as

$$\omega(\dot{\gamma}(t)) = dH.$$

Although we have only touched upon applications in physics and differential geometry, we see that this bridge can be crossed [1].

Chapter 4

Conclusion

The Calculus of Variations is a highly applicable and advancing field. My thesis has only scraped the top of the applications and theoretical work that is possible within this branch of mathematics. To summarize, we began by exploring a general problem common to this field, finding the geodesic between two given points. We then went on to define and explore terms and concepts needed to further delve into the subject matter. In Chapter 2, we examined a special set of smooth functions, inspired by the Calabi extremal metric, and used some general theory of convex functions in order to determine the minimizer of the functions. Finally, we looked into Hamilton's equations and touched upon symplectic geometry.

For those people looking to investigate further into the calculus of variations, there are many different paths to take. A particularly interesting one would be to explore the specific applications the field has to other disciplines, like physics and economics. We briefly explored the applications in physics, but did not explore applications in economics at all. One could learn about the problems of exhaustible resources, which demand knowledge of the calculus of variations. In this specific application, the extremum that we are looking for is the maximal profit [4].

This branch of mathematics is useful to a variety of other disciplines, and it would be beneficial to explore how the fields coincide in order to solve real-world problems. We have explored some theoretical mathematics of this field, so this would be the natural next step. There is also the possibility that applications for the Calculus of Variations have not yet been realized, leaving the potential for a large amount of growth and progress.

Bibliography

[1] Berndt, Rolf. "Some Aspects of Theoretical Mechanics." An Introduction to Symplectic Geometry. Providence, RI: American Mathematical Society, 2001. 1-5. Print.

[2] Calabi, Eugenio. *Extremal Kahler Metrics*. Seminar on Differential Geometry, 259-290, Princeton University Press, Princeton, NJ, 1982.

[3] Goldstein, Herbert. "Examples of Canonical Transformations." Classical Mechanics. Cambridge, MA: Addison-Wesley, 1980. 389-90. Print.

[4] Hotelling, Harold. "The Economics of Exhaustible Resources." Journal of Political Economy 39.2 (1931): 137-75. Print.

[5] Pressley, Andrew. *Elementary Differential Geometry.* London: Springer, 2001. Online.

[6] Van, Brunt B. *The Calculus of Variations.* New York: Springer, 2004. Print.