Dipartimento di Impresa e Management
Management and Computer Science

Calculus of Variations: Euler-Lagrange equation and Hamiltonian formulation

SUPERVISOR: Prof. Giovanni Alessandro Zanco

CANDIDATE: Alessio Francis Denny (245331)

ACADEMIC YEAR: 2021/2022

Acknowledgments

A sincere acknowledgement goes to my family's support.

I am also grateful to professor Zanco for his patience, expertise, continuous commitment and dedication in helping me improve upon the thesis, which has made this project a great learning experience for me. I also want to thank professor Zanco for always being available to meet and discuss how the thesis is progressing, as well as for his continuous feedback.

Table of contents

Aim of the Thesis

The structure of the thesis will be in the following manner. A general introduction to differential equations will be given including classification of differential equations, some examples of differential equations, as well as the most important theorems/proofs connected to differential equations. Furthermore, the intuition behind Cauchy sequences, Cauchy sequences of functions, uniform and pointwise convergence of sequences of functions will be thoroughly explained. Next, the intuition behind the concept of optimization will be given. The Euler-Lagrange differential equation will then be thoroughly investigated and will be applied to solve the infamous Brachistochrone problem, which is an example of a control theory problem. Additionally, this thesis will explore some special cases of the Euler-Lagrange differential equation as well as integral and non-integral constraints.The Hamiltonian will also be discussed and its physical interpretation will be analyzed. Lastly, the thesis will provide some concluding remarks, concerning some applications of the Euler-Lagrange differential equation to Computer Science, for example Machine Learning.

Chapters 1 and 6 are based on the book “Mathematics for Economists”, written by Carl P Simon and Lawrence E.Blume, while chapters 7 through 12 are based on the book “ Calculus of Variations and Optimal Control Theory: A Concise Introduction” written by Daniel Liberzon. All the graphs and visual representations included in this thesis were realized by the author using the Geogebra software as well as other open-source graphing software.

PART 1

Chapter 1: Background theory of differential equations

In its most basic form, a differential equation is a relationship between a function, which can be of one variable or of more than one variable, and its derivatives. Once we solve a differential equation, the solution will be a function. In the first case, when dealing with differential equations, we speak of Ordinary Differential Equations, and in the latter case, we speak of Partial Differential equations. This thesis will examine Ordinary Differential Equations.

Differential equations are powerful mathematical tools that can be used to model real life phenomena. Differential equations have a wide variety of applications throughout different fields of study. More specifically, they have applications in Physics, Chemistry, Economics, Epidemiology, etc. As an example, consider the following differential equation

$$y'(t) = \alpha \cdot y(t) \quad \text{where } \alpha \in \mathbb{R}$$

The differential equation describes a situation in which the derivative of the quantity y at time t is proportional to the quantity itself.

The general form of an Ordinary Differential equation is given by

$$f\left(t,\ y(t),\ y'(t),\ \ldots,\ y^{(n-1)}(t),\ y^{(n)}(t)\right) = 0$$

Where t is the independent variable. Moreover, let's define

$y: \mathrm{Domain}(y) \subset \mathbb{R} \to \mathbb{R}$ as being the solution to the differential equation.

$y^{(j)}$, where $j = 1, 2, 3, \ldots, n$ are the derivatives of j'th order.

$f: \mathrm{domain}(f) \subset \mathbb{R}^{n+2} \to \mathbb{R}$ where the number n is the order of the ordinary differential

Now consider the following expression:

$$y^{(n)}(t) = f\left(t,\ y(t),\ y'(t),\ y''(t),\ \ldots,\ y^{(n-1)}(t)\right)$$

$f: \mathrm{Domain}(f) \subset \mathbb{R}^{n+1} \to \mathbb{R}$

If an Ordinary Differential Equation can be written as in the expression above, then we say that the Ordinary Differential Equation is in normal form.

The ordinary differential equation can be transformed in a more convenient way as follows:
Define:

$$y_1(t) = y(t)$$

$$y_2(t) = y'(t)$$

$$y_3(t) = y''(t)$$

$$\vdots$$

$$y_n(t) = y^{(n-1)}(t)$$

$$y_{n+1}(t) = y^{(n)}(t)$$

Let's pay close attention to the following equation:

$$y_{n+1}(t) = y^{(n)}(t)$$

The equation above is equal to the original differential equation, namely:

$$y_{n+1}(t) = y^{(n)}(t) = f\left(t, y(t), y'(t), y''(t), \ldots, y^{(n-1)}(t)\right)$$

$$y_{n+1}(t) = y^{(n)}(t) = f(t, y_1(t), y_2(t), \ldots, y_n(t))$$

Moreover, notice that

for $j = 1, \ldots, n$ we have $y_j(t) = y^{(j-1)}(t) \Rightarrow y'_j(t) = y^{(j)}(t)$

Hence, we have obtained that

$$y_{n+1}(t) = y^{(n)}(t) = f(t, y_1(t), y_2(t), \ldots, y_n(t)) = y_n'(t)$$

If we denote

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$ and $\vec{F}$ as being a vector function, we can express the original differential equation as follows:

$$y'(t) = \vec{F}(y(t))$$

This representation is very useful because it shows an Ordinary Differential equation of any order written in normal form can be reduced to a system of first order Ordinary Differential Equations, which in most cases is easier to solve.

Now that we have introduced the most general form of a differential equation in $\mathbb{R}^n$, let's define the different types of differential equations, including some simple examples of each category as well.

Differential equations can be autonomous or non-autonomous, first order or higher order, linear, non-linear, homogeneous, non-homogeneous.

An autonomous differential equation is a differential equation of the form

$$f\left(t,\ y(t),\ y'(t),\ \ldots\ ,\ y^{(n-1)}(t),\ y^{(n)}(t)\right) = 0$$

whose left hand side is independent of the variable t. The equation below is an example of a first order autonomous differential equation.

$$y'(t) = f(y)$$

EXAMPLE 1 (autonomous differential equation)

Consider the ODE given by $y' = y$

To solve it we can use the separation of variables technique, since the differential equation is separable. An Ordinary Differential Equation is separable if it can be written in the form

$y'(t) = f(t, y)$

The right hand side namely $f(t, y)$ must be factored in the form $f(t, y) = g(t) \cdot h(y)$ where g and h are known functions.

In this case, we obtain:

$$\frac{dy}{dt} = y \qquad h(y) = y \text{ and } g(t) = 1$$

$$dy = y \cdot dt$$

$$\frac{1}{y}\, dy = dt$$

In order to get rid of the derivative operators, we take the indefinite integral of both sides of the expression. We are implicitly using the Fundamental Theorem of Calculus that states that integration is the inverse operation to differentiation:

$$\int \frac{1}{y}\, dy = \int dt$$

$$\ln(y) = t + c \quad , \text{ where } c \in \mathbb{R}$$

To solve for the variable “y”, we take the inverse of the natural logarithm of both sides:

$$e^{\ln(y)} = e^{t+c}$$

$$y = e^{t} \cdot e^{c}$$

Since e^c is another constant, let's denote e^c as C_1, in which $C_1 \in \mathbb{R}$

$$y = C_1 \cdot e^{t}$$

The solution to the differential equation is just a scaled version of the exponential function $y = e^t$.

So far, we have obtained general solutions to differential equations. If we want to obtain a particular solution to a differential equation, we need to specify some initial condition. This is called a Cauchy problem. More specifically,

$$\begin{cases} y'(t) = f(t, y) \\ y(t_0) = x_0 \end{cases}$$

defines a Cauchy problem. We use the initial value condition to figure out a particular solution to the ODE.

Non-autonomous differential equations are different from autonomous differential equations since the right hand side of the differential equation can depend both on y and t. Let's consider a more challenging example:

EXAMPLE 2 (non − autonomous differential equation)

Consider the following differential equation:

$$\begin{cases} \frac{dy}{dt} = 4 + \left(\frac{y}{t}\right)^2 - \left(\frac{y}{t}\right) \\ y(1) = 2 \end{cases}$$

The above differential equation is non-autonomous because the right hand side depends explicitly on t.

To solve this differential equation, we can apply a simple substitution, namely:

$$\text{Let } \psi = \frac{y}{t} \quad \Rightarrow y = t \cdot \psi$$

Next, we can differentiate the above expression with respect to t:

$$\frac{dy}{dt} = \psi + t \cdot \frac{d\psi}{dt}$$

Next, let's substitute into the original differential equation:

$$\psi + t \cdot \frac{d\psi}{dt} = 4 + \psi^2 - \psi$$

Since the above differential equation is separable, we can algebraically manipulate it to have each one of the variables on one side:

$$t \cdot \frac{d\psi}{dt} = 4 + \psi^2 - 2\psi$$

$$\frac{1}{\psi^2 - 2\psi + 4} d\psi = \frac{1}{t}\, dt$$

Notice that $\psi^2 - 2\psi + 4 \neq 0$ when dividing by zero. This implies that we may not be able to solve the ODE for $t = 0$. Nevertheless, we may expect to find some interval around the inital $t = 1$ on which the ODE can be solved.

Like in the previous example, we can take the indefinite integral of both sides to get rid of the differentials:

$$\int \frac{1}{\psi^2 - 2\psi + 4} d\psi = \int \frac{1}{t}\, dt \qquad \text{(Equation 1)}$$

The integral on the right hand side is immediate to solve. However, the integral on the left hand side is not as straightforward. Let's first solve the integral on the left hand side and then we can put all the pieces together to solve the differential equation.

We proceed by writing the denominator of the fraction in vertex intercept form:

$$\int \frac{1}{\psi^2 - 2\psi + 4} d\psi = \int \frac{1}{(\psi - 1)^2 + 3} d\psi$$

We can now apply the substitution:

$$\int \frac{1}{(\psi-1)^2+\sqrt{3}^2}\,d\psi \qquad \text{let } u = (\psi-1),\ \frac{du}{d\psi} = 1$$

$$\int \frac{1}{u^2+\sqrt{3}^2}\,du$$

(Equation 1.1)

We will now apply a trigonometric substitution:

$$\theta = \arctan\left(\frac{u}{\sqrt{3}}\right) \text{ so that } u = \sqrt{3}\cdot\tan(\theta),\ \frac{du}{d\theta} = \sqrt{3}\cdot\sec^2(\theta)$$

Then, $\displaystyle\int \frac{1}{u^2+\sqrt{3}^2}\,du =$

$$= \int \frac{1}{3\cdot\tan^2(\theta)+3}\cdot\sqrt{3}\cdot\sec^2(\theta)\,d\theta$$

$$= \frac{1}{3}\int \frac{1}{1+\tan^2(\theta)}\cdot\sqrt{3}\cdot\sec^2(\theta)\,d\theta$$

Recall that:

$$1+\tan^2(\theta) = \sec^2(\theta)$$

Hence, the integral above simplifies to:

$$\frac{1}{3}\int \frac{1}{\sec^2(\theta)}\cdot\sqrt{3}\cdot\sec^2(\theta)\,d\theta =$$

$$= \frac{\sqrt{3}}{3}\int d\theta$$

$$= \frac{1}{\sqrt{3}}\cdot\theta + c_1 \quad \text{with } c_1 \in \mathbb{R}$$

Substituting back we obtain:

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{u}{\sqrt{3}}\right) + c_1$$

And since

$$u = \psi - 1$$

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{\psi - 1}{\sqrt{3}}\right) + c_1$$

We can figure out the solution to the differential equation:

Equation 1 is equivalent to:

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{\psi - 1}{\sqrt{3}}\right) + c_1 = \ln(t) + c_2 \qquad \text{such that } c_1,\ c_2 \in \mathbb{R}$$

(equation 1.2)

Which can be written as

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{\psi - 1}{\sqrt{3}}\right) = \ln(t) + C \quad \text{with } C \in \mathbb{R}$$

Let's now figure out the value of the constant C using the initial value condition:

Recall that $\psi = \dfrac{y}{t}$ and that $y(1) = 2$, thus

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{2-1}{\sqrt{3}}\right) = \ln(1) + C$$
$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{1}{\sqrt{3}}\right) = C$$

$$\frac{\pi}{6 \cdot \sqrt{3}} = C$$

Finally, plugging the value of the constant C into **equation 1.2**, we obtain the solution to the non-autonomous differential equation:

$$\frac{1}{\sqrt{3}} \cdot \arctan\left(\frac{\frac{y(t)}{t} - 1}{\sqrt{3}}\right) = \ln(t) + \frac{\pi}{6\sqrt{3}}$$

$$\arctan\left(\frac{\frac{y(t)}{t} - 1}{\sqrt{3}}\right) = \sqrt{3} \cdot \ln(t) + \frac{\pi}{\sqrt{6}}$$

$$\frac{\frac{y(t)}{t} - 1}{\sqrt{3}} = \tan\left(\sqrt{3} \cdot \ln(t) + \frac{\pi}{\sqrt{6}}\right)$$

$$\frac{y(t)}{t} - 1 = \sqrt{3} \cdot \tan\left(\sqrt{3} \cdot \ln(t) + \frac{\pi}{\sqrt{6}}\right)$$

$$\frac{y(t)}{t} = 1 + \sqrt{3} \cdot \tan\left(\sqrt{3} \cdot \ln(t) + \frac{\pi}{\sqrt{6}}\right)$$

$$y(t) = t \cdot \left[1 + \sqrt{3} \cdot \tan\left(\sqrt{3} \cdot \ln(t) + \frac{\pi}{\sqrt{6}}\right)\right]$$

Anyways, linear ODEs deserve particular attention because they can often be solved explicitly.

A first order linear differential equation takes the form:

$$a_0(t)\cdot y(t) + a_1(t)\cdot y'(t) + \ldots + a_n(t)\cdot y^{(n)}(t) = b(t)$$

It is Homogeneous if

$$b(t) = 0 \ , \ \forall t$$

And non-homogeneous otherwise. Additionally, It is said to have constant coefficients if all the terms $a_0, a_1, \ldots, a_n$ are constants and do not depend on t.
We already saw a First Order homogeneous linear differential equation with constant coefficients in example 1.

A second order linear homogeneous differential equation with constant coefficients takes the form:

$$a\cdot \frac{d^2y}{dt^2} + b\cdot \frac{dy}{dt} + cy = 0 \quad \text{with } a, b, c \in \mathbb{R}$$

To solve a second order linear homogeneous differential equation, we can proceed in the following manner:

Assume as in the first order case that $y = e^{rt}$ is a solution. Then, $\frac{dy}{dt} = re^{rt}$ and $\frac{d^2y}{dt^2} = r^2e^{rt}$

Plugging these values into the differential equation we obtain:

$$a\cdot r^2\cdot e^{rt} + br\cdot e^{rt} + c\cdot e^{rt} = 0$$

$$e^{rt}\cdot\left(ar^2 + br + c\right) = 0$$

Since e^{rt} is always different than zero, this implies that $ar^2 + br + c = 0$

The equation

$$ar^2 + br + c = 0$$

Is called the characteristic equation of the differential equation. Dependending on the discriminant of the quadratic polynomial on the left hand side, we can understand the nature of the roots of the equation:

The ODE has a solution in all three cases and it takes different forms. An example of the last case will be provided below.

If $b^2 - 4ac < 0$ there are two complex roots

If $b^2 - 4ac = 0$ there is one multiple root

If $b^2 - 4ac > 0$ there are two real and distinct roots

Let's consider an example:

EXAMPLE 3 (second – order linear homogeneous differential equation with constant coefficients)

$$\frac{dy^2}{dt^2} - \frac{dy}{dt} - 2y = 0$$

$r^2 - r - 2 = 0$ is the characteristic equation

The discriminant of the characteristic equation is given by:

$1 - 4 \cdot (-2) = 9 > 0$

Hence, the characteristic equation has two real roots

The characteristic equation can be factored in the following way:

$$r^2 - r - 2 = 0$$
$$(r - 2) \cdot (r + 1) = 0$$

Then, the two roots are:

$r = 2$, $r = -1$

Hence, the general solution takes form:

$$y = C_1 \cdot e^{2t} + C_2 \cdot e^{-t} \text{ with } C_1, C_2 \in \mathbb{R}$$

The last class of differential equations that will be discussed in this thesis is the non-homogeneous second order differential equations with constant coefficients. The most general form of a non-homogeneous second order differential equation with constant coefficients is:

$$a \cdot \frac{d^2y}{dt^2} + b \cdot \frac{dy}{dt} + c \cdot y = g(t) \text{ with } a, b, c \in \mathbb{R} \text{ and } g: \text{Domain}(g) \subset \mathbb{R} \to \mathbb{R}$$

The term $g(t)$ is called the "forcing term" . It turns out that there is a theorem that explains how to construct the solution to the non-homogeneous second order differential equations with constant coefficients; it will not be proved in this thesis. The statement of the theorem given here is based on Theorem 24.4 from the book Mathematics for economists by Carl P Simon and Lawrence E.Blume.

THEOREM (non – homogeneous differential equation with constant coefficients)

Let $y_p(t)$ represent a particular solution to the non – homogeneous differential equation (3) .

Moreover, let $y_0(t)$ be the general solution to the homogeneous differential equation

$a \cdot \frac{d^2y}{dt^2} + b \cdot \frac{dy}{dt} + c \cdot y = 0$ with $a, b, c \in \mathbb{R}$.

Then, the general solution to the non – homogeneous differential equation is given by

$y(t) = y_0(t) + y_p(t)$

The theorem above is powerful since it is saying that we only need to find a particular solution to the non-homogeneous differential equation and the general solution to the homogeneous one, which is an easy take. To find a particular solution, the method of undetermined coefficients is usually employed. Let's consider a simple example:

EXAMPLE 4 (second order non – homogeneous differential equation)

Solve $\frac{d^2y}{dt^2} - 2 \cdot \frac{dy}{dt} - 3y = 9t^2$

We begin by solving the homogeneous form of this differential equation. Using the same method as in example 3, the solution to the homogeneous differential equation is given by:

$$y(t) = C_1 \cdot e^{3t} + C_2 \cdot e^{-t}, \text{ such that } C_1, C_2 \in \mathbb{R}$$

Now, we need to find a particular solution. We proceed in the following manner:

Define $y_p = At^2 + Bt + C$ such that $A, B, C \in \mathbb{R}$ and y_p denotes a particular solution. Then,

$$\frac{dy_p}{dt} = 2At + B$$

$$\frac{d^2y_p}{dt^2} = 2A$$

After we substitute the expressions that we found above into the original differential equation, we obtain:

$$2A - 2 \cdot (2At + B) - 3 \cdot \left(At^2 + Bt + C\right) = 9t^2$$

$$2A - 4At - 2B - 3At^2 - 3Bt - 3C = 9t^2$$

$$2A - 2B + t \cdot (-4A - 3B) + t^2 \cdot (-3A) - 3C = 9t^2$$

We can now match the terms on the right and left hand side:

$$-3A = 9 \rightarrow A = -3$$

$$-4A - 3B = 0 \rightarrow B = 4$$

$$2A - 2B - 3C = 0 \rightarrow C = -\frac{14}{3}$$

Hence, a particular solution is given by:

$$y_p = -3t^2 + 4t - \frac{14}{3}$$

The general solution of the nonhomogeneous ODE is given by:

$$y(t) = C_1 \cdot e^{3t} + C_2 \cdot e^{-t} - 3t^2 + 4t - \frac{14}{3}$$

Chapter 2: Picard-Lindelöf general statement of the Theorem

In this section of the thesis, we will explore what the Picard-Lindelöf theorem states, as well as the significance of this theorem. The statement of the theorem is the following:

THEOREM

If $f : \left[t_0 - a,\, t_0 + a\right] \times \left[y_0 - b\,,\, y_0 + b\right] \to \mathbb{R}$ *is continuous and such that*
$\forall (t,y), (t,z) \in \left[t_0 - a,\, t_0 + a\right] \times \left[y_0 - b\,,\, y_0 + b\right]$ *we have* $\left|f(t,y) - f(t,z)\right| \le L \cdot \left|y - z\right| \; \forall\, t \in \left[t_0 - a,\, t_0 + a\right]$

then $\exists\ \tilde{a}$ *such that* $\begin{cases} y'(t) = f(t,\, y(t)) \\ y(t_0) = y_0 \end{cases}$ *has a unique solution in the interval* $\left[t_0 - \tilde{a}\ ,\ t_0 + \tilde{a}\right]$

The first part of the Picard-Lindelöf theorem, namely

$f : \left[t_0 - a,\, t_0 + a\right] \times \left[y_0 - b\,,\, y_0 + b\right] \to \mathbb{R}$ *is continuous and such that*
$\forall (t,y), (t,z) \in \left[t_0 - a,\, t_0 + a\right] \times \left[y_0 - b\,,\, y_0 + b\right]$ *we have* $\left|f(t,y) - f(t,z)\right| \le L \cdot \left|y - z\right| \; \forall\, t \in \left[t_0 - a,\, t_0 + a\right]$

Is essentially the definition of a Lipschitz function. Let's consider an example of a Lipschitz function:

EXAMPLE **(Lipschitz Function)**

Consider $D = \mathbb{R}^2$ *and define* $f(t,y) = t^2 + 2y$ *. For every* $(t,y)\,,\ (t,z) \in D$ consider

$$\left|f(t,y) - f(t,z)\right| = \left|t^2 + 2y - \left(t^2 + 2z\right)\right| = = \left|2y - 2z\right| = 2 \cdot \left|y - z\right|$$

Hence, the function f satisfies the Lipschitz condition on D with $L = 2$

The Picard-Lindelöf theorem is important since it provides a specific condition under which we are sure to find a unique solution to our Cauchy problem. Even though when we will explore the Brachistochrone problem, we will consider a curve that is fixed at two points, and the Picard-Lindelöf theorem considers only one boundary condition, the Picard-Lindelöf theorem is still an important pillar in Calculus of Variations.

The complete proof of this theorem will be given in a later chapter of this thesis. To prove the Picard-Lindelöf theorem, several mathematical tools are now introduced, beginning with Cauchy Sequences, Cauchy Sequences of functions, as well as pointwise and uniform convergence of sequences of functions.

Chapter 3: Intuition behind Cauchy sequences

Now that we have introduced the basic background theory behind differential equations, we are ready to move on to exploring the Picard-Lindelof theorem which provides a condition that guarantees that there exists a solution to a Cauchy problem. However, before exploring the Picard-Lindelof theorem and the Euler-Lagrange differential equation, it is imperative to understand the concept of Cauchy sequences. Let's begin by exploring the concept of a Cauchy sequence.

Mathematically, a sequence $\{x_n\}_{n \in N \subset R}$ is a Cauchy sequence if by definition:

DEFINITION

$\forall\ \varepsilon > 0$, $\exists$ N such that $\forall$ m, n $\geq$ N , $\left| x_m - x_n \right| < \varepsilon$

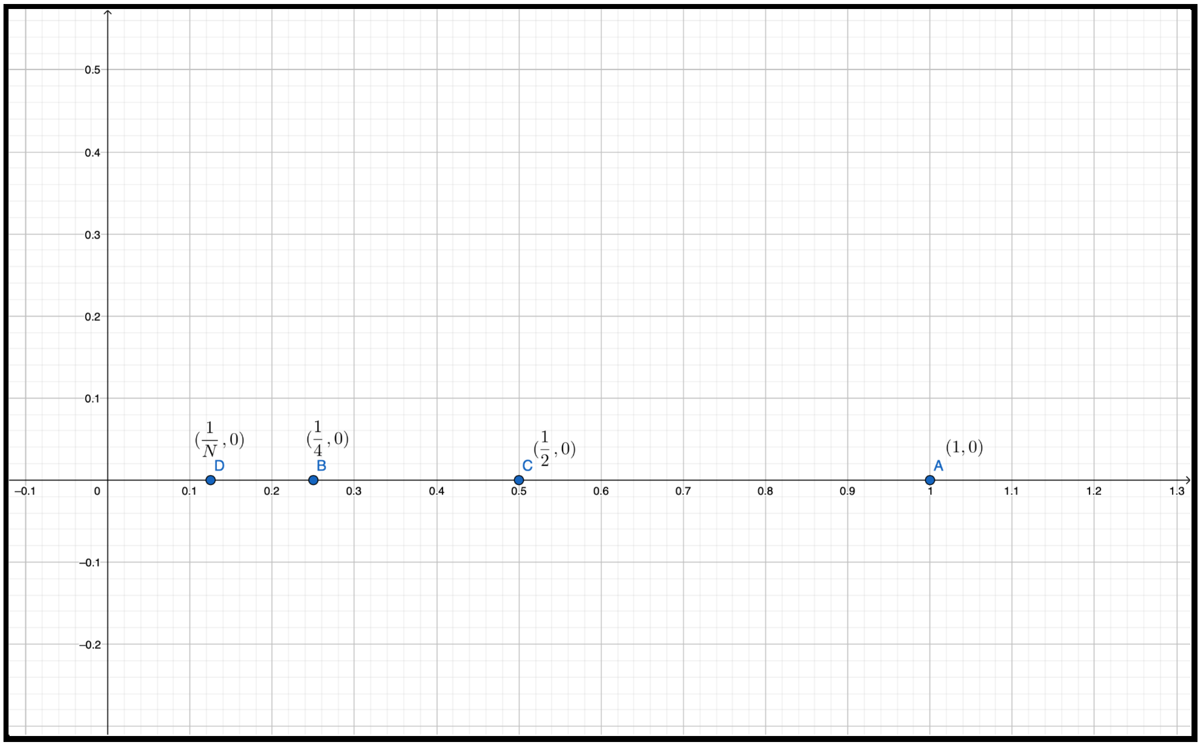
Let's try to gain some more intuition behind a Cauchy sequence.

Let's consider an example of a Cauchy sequence. Consider the metric space defined by $(\mathbb{R},\ d)$ in which d denotes the Euclidean metric. Moreover, consider the following sequence:

$$x: \mathbb{N} \rightarrow \mathbb{R}$$

$$x_n = \frac{1}{n}$$

The sequence maps each input to its reciprocal. **Figure 1** provides a visual representation of this sequence:

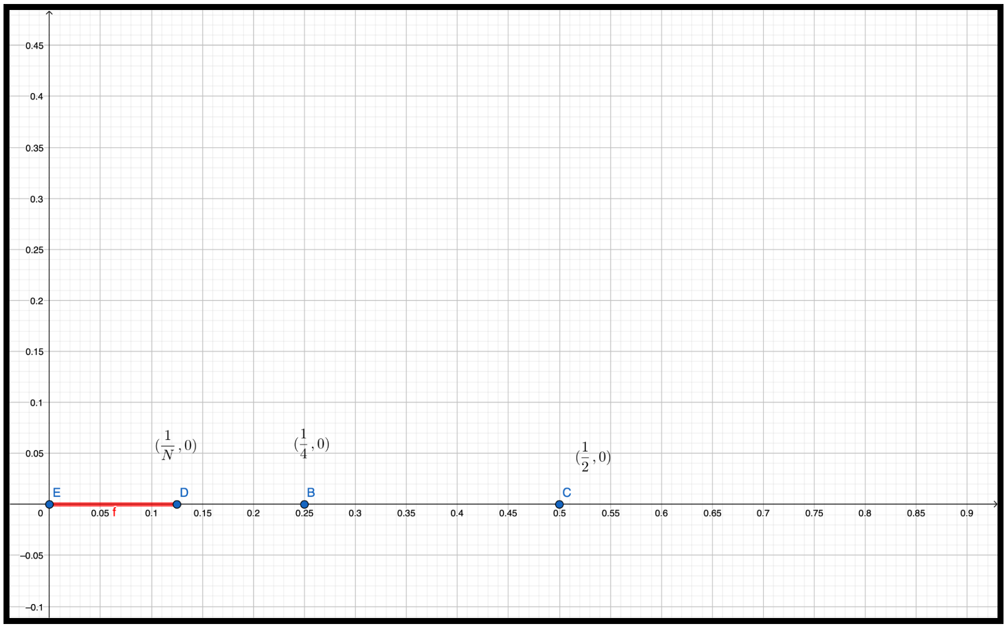


(Figure 1)

Notice that the sequence seems to converge to zero, since the terms keep on getting closer to zero. Let's show that this sequence is Cauchy. Let $\epsilon > 0$. We want to find an N such that after this point the distance between any two point is less than ϵ.

If we pick an arbitrary point denoted by $\frac{1}{N}$, *we know that the distance between any two points after* $\frac{1}{N}$ *cannot be greater than zero, since* 0 *is the infimum of the sequence.*

Figure 2 provides a visual representation. The red line represents the distance.



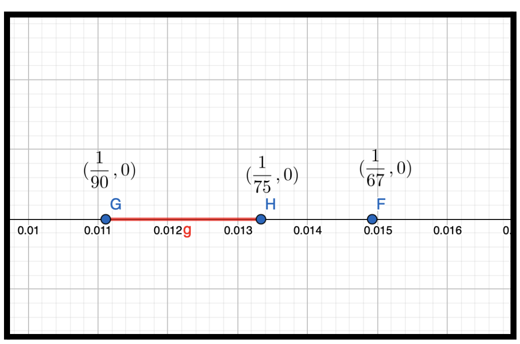
(Figure 2)

This means that

$$d\left(0, \frac{1}{N}\right) < \varepsilon$$

$$\left|\frac{1}{N}\right| < \varepsilon \quad \rightarrow N > \frac{1}{\varepsilon}$$

Hence, if we pick for example $\epsilon = 0.015$, this means that $N > 66.6667$. If we pick $N = 67$, this means that beyond the terms $\frac{1}{67}$, the distance between any two numbers must be less than $\epsilon = 0.015$. For example, $d(\frac{1}{90}, \frac{1}{70}) = |-0.003175| = 0.003175 < 0.015$. **Figure 3** provides some more visual support. The value $d(\frac{1}{90}, \frac{1}{70})$ is depicted by the red line.



(Figure 3)

Additionally, Cauchy sequences of numbers have a very interesting property. The theorem states that if a sequence is Cauchy, this implies that it is convergent. The proof will be discussed below.

The first part of the proof will be to show that every sequence has a monotone subsequence.

By definition, a number $N \in \mathbb{N}$ *is a peak for* $\{x_n\}$ *if* $m > N$ *implies that* $x_m < x_n$.

What this is saying is that any term of the sequence after the peak term of the sequence must be smaller than the peak term.

Given a sequence $\{x_n\}$ *, either it has infinitely many peaks or finitely many peaks. If there are infinitely many peaks denoted by* $N_1, N_2, N_3, \ldots$ *this implies that the sequence* $x_{N_1}, x_{N_2}, x_{N_3}, \ldots$ *is a monotone subsequence.*

On the other hand, if the sequence $\{x_n\}$ *has a finite number of peaks, then there must be the last peak; denote it by* N_k.

More specifically, $\exists\ n_1 > N_k$ *such that* $x_{n_1} < x_{N_k}$. *Hence,* n_1 *is not a peak.*
This further implies that $\exists\ n_2 > n_1$ *such that* $x_{n_2} > x_{n_1}$.

We can iterate this process to find that $x_{n_1} < x_{n_2} < x_{n_3} < \ldots$
The sequence above is an increasing subsequence.

The second part of the proof will be to show that a Cauchy sequence is necessarily bounded.

Let $\{x_n\}$ *be a Cauchy sequence and let* $\varepsilon = 1$. *Then,* $\exists\ N$ *such that* $\forall\, m, n \geq N\ |x_m - x_n| < 1$.
In particular, $\forall\, n \geq N\ |x_n - x_N| < 1$. *Moreover, denote by* $M = max\left\{ |x_1|, |x_2|, \ldots, |x_N|, |x_N| + 1 \right\}$;
if $n \leq N$, *we can state that* $|x_n| \leq M \leq M + 1$

If $n > N$, *then* $|x_n| = |x_n - x_N + x_N| \leq |x_n - x_N| + |x_N|$ *using the Triangle Inequality*

By assumption that $\{x_n\}$ *is a Cauchy sequence, we know that* $|x_n - x_N| < 1$. *From before, we also know that* $|x_N| \leq M \; \forall \, n$

Hence, this implies that $|x_n| \leq M, \; \forall \, n \geq 0$, *thus the sequence* $\{x_n\}$ *is bounded.*

So far, we have shown that a Cauchy sequence is bounded, which further implies that all its monotone subsequences are convergent (we have proved above that it must have at least one convergent subsequence). Moreover, a sequence being Cauchy and having a convergent subsequence implies that the entire sequence is convergent. Let's prove this last statement:

We can affirm that if a sequence $\{x_n\}_n$ *is Cauchy, it has a convergent subsequence* $\left\{x_{n_j}\right\}_j$.

Let us denote by x its limit, so that $\lim\limits_{j \to \infty} x_{n_j} = x$.

It is important to first fix $\varepsilon > 0$.

Since the sequence $\left\{x_{n_j}\right\}_j$ converges to x, this implies by definition that $\exists \; J$, such that $\forall \; j \geq J$

$$\left|x_{n_j} - x\right| < \frac{\varepsilon}{2}.$$

Additionally, since $\left\{x_n\right\}_n$ is a Cauchy sequence, we know that $\exists \; \tilde{N}$, such that $\forall \; m, n \geq \tilde{N}$

$$\left|x_m - x_n\right| < \frac{\varepsilon}{2}$$

Next, $\forall \; n \geq \max\left\{\tilde{N}, n_J\right\}$, $\left|x_n - x\right| \leq \left|x_n - x_J\right| + \left|x_{n_J} - x\right| < \dfrac{\varepsilon}{2} + \dfrac{\varepsilon}{2} = \varepsilon$

Finally, this implies that the original sequence $\left\{x_n\right\}$ converges to x.

Chapter 4: Intuition behind Cauchy sequences of functions, pointwise convergence and uniform convergence

In the previous section of this thesis, we explored the concept of a Cauchy sequence. In this section, we will investigate what it means for a sequence of functions to be Cauchy, and we will discuss how to show that a Cauchy sequence of continuous functions is convergent.

$Let's\ define\ I\ as\ being\ an\ interval\ \ and\ \mathbb{C} = \mathbb{C}(I, \mathbb{R}) = \{continuous\ functions\ f: I \mapsto \mathbb{R}\}$

$Moreover,\ consider\ \ a\ sequence\ \{f_n\}_{n \in \mathbb{N}} \subset \mathbb{C},\ that\ is\ a\ set\ of\ functions\ indexed\ by\ n\ such\ \ that\ f_n \in \mathbb{C}(I, \mathbb{R})\ \forall\, n$

DEFINITION

$\{f_n\}_{n \in \mathbb{N}}$ $converges\ pointwise\ to\ a\ function\ f\ \ if\ f_n(t) \to f(t)\ \forall\, t \in I$

As a concrete example, consider the function

$$\mathrm{f_n(t)} = 2\mathrm{t} + \frac{\mathrm{t}}{\mathrm{n}^2}$$

Notice that as the variable n approaches infinity, we obtain:

$f_n(t) \xrightarrow{n \to \infty} 2t\ ,\ \forall\, t \in I$

$Therefore,\ f_n(t)\ converges\ pointwise\ to\ f(t) = 2t\ ,\ \forall\, t \in I$

Furthermore, let’s consider another important definition:

DEFINITION

$A\ \ sequence\ of\ functions\ \{f_n\} \subset \mathbb{C},\ converges\ uniformly\ to\ a\ function\ f\ \ if\ \ \sup_{t \in I} \left| f_n(t) - f(t) \right| \to 0$

What the statement above is saying is that a sequence of functions that are continuous and that belong to the interval converge to a function if the supremum of the absolute value of the difference between each of the functions and the converging function goes to zero.

Let’s consider more closely the expression

$$\sup_{t \in I} \left| f_n(t) - f(t) \right| \to 0$$

The norm of a continuous function is defined as:

$\|f\|_\infty = \sup_{t \in I} |f(t)|$, *therefore the quantity* $d(f,g) = \|f - g\|_\infty = \sup_{t \in I} \left| f(t) - g(t) \right|$ *which is a distance on* $\mathbb{C}$.

Now, we are ready to define what a Cauchy sequence of functions is.

A sequence $\{f_n\}$ *of continuous functions is a Cauchy sequence if* $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$, *such that* $\forall m, n \geq N$, $\|f_m - f_n\|_\infty < \varepsilon$. *In other words the norm of* $f_m - f_n$ *must be less than* ε *provided m and n are big enough.*

The term $\|f_m - f_n\|_\infty < \varepsilon$ *can be re – written as* $d(f_m - f_n) < \varepsilon$.

Before moving on to an important proof, let's consider some examples of sequences of functions that are pointwise and/or uniform convergent to clarify these ideas.

EXAMPLE 5 (pointwise and uniform convergence)

Discuss the pointwise and uniform convergence of $f_n(x) = n \cdot e^{-(nx)^2}$, $\forall x \in \mathbb{R}$

We begin by checking whether or not the sequence above is pointwise convergent.

We can rewrite the sequence of functions in the following form:

$$f_n(x) = \frac{n}{e^{(nx)^2}}$$

We will now compute the $\lim_{n \to \infty} f_n(x)$ considering two different cases:

$$\lim_{n \to \infty} f_n(x) = \begin{cases} 0, & if\ x \neq 0 \\ +\infty, & if\ x = 0 \end{cases}$$

Hence, we have figured out that $f_n(0) = n$ if $x = 0$ and the $\lim_{n\to\infty} f_n(0) = +\infty$. Thus, the sequence $f_n(x)$ does not converge pointwise for every $x \in \mathbb{R}$ since it diverges to $+\infty$ if $x = 0$.

Since the sequence of functions $f_n(x)$ does not converge pointwise for every $x \in \mathbb{R}$, we know that it does not converge uniformly on $\mathbb{R}$.

For example, consider the following expression

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}\setminus\{0\}} \left|\frac{n}{e^{(nx)^2}} - f(x)\right| = \lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\frac{n}{e^{(nx)^2}} - 0\right| = \lim_{n\to\infty} |n| = +\infty \quad \textit{if } x \neq 0$$

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\frac{n}{e^{(nx)^2}} - \infty\right| = \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |n - \infty| \quad \textit{which is undefined if } x = 0$$

The sequence does not converge uniformly on $\mathbb{R}$.

Let's consider another example:

EXAMPLE 6 (pointwise and uniform convergence)

Discuss the pointwise and uniform convergence of $f_n(x) = \begin{cases} 1, & \textit{for } x \in [n, n+1) \\ 0, & \textit{otherwise} \end{cases}$

We begin by investigating the pointwise convergence of $f_n(x)$ like in the previous example. We begin by considering an arbitrary $x \in \mathbb{R}$. This means that $\exists\, n_0 \in \mathbb{N}$ such that $x \in [n_0, n_0 + 1]$. Thus, $\forall\, n \geq n_0 + 1$ we know that $f_n(x) = 0$. Hence, we can state that $\lim_{n\to\infty} f_n(x) = 0 = f_0(x)$. Hence, this proves that $f_n(x)$ converges pointwise to $f_0(x)$ on $\mathbb{R}$.

We will now investigate the uniform convergence of the function $f_n(x)$ in $\mathbb{R}$. We will begin by assuming that $f_n(x)$ converges uniformly to $f_0(x)$ on $\mathbb{R}$. Then, given that $\epsilon = \frac{1}{2}$, $\exists\ a\ unique$ $n_0 \in \mathbb{N}$, such that for every $n \geq n_0$ and every $x \in \mathbb{R}$, we have that

$|f_n(x)$-$f_0(x)| = f_n(x) < \frac{1}{2}$.

Now, let's take an arbitrary $n \geq n_0$ and an $x \in [n, n+1)$. This implies that

$|f_n(x)$-$f_0(x)| = f_n(x) = 1 < \frac{1}{2}$. This is a contradiction, which proves that the sequence of functions $f_n(x)$ does not converge uniformly on the set $\mathbb{R}$.

Lastly, in this section of the thesis, we will prove the following theorem

THEOREM

A Cauchy set of continuous functions implies that they are uniformly convergent

PROOF

We begin by fixing $\varepsilon > 0$. *Find a* N *such that* $\forall\ m, n \geq N$, $\left\|f_m - f_n\right\|_\infty < \varepsilon$

Next, we fix a $t \in I$. *This implies that* $\left|f_m(t) - f_n(t)\right| \leq \sup_{s \in I}\left|f_m(s) - f_n(s)\right|$.

The reason that the left hand side is less than or equal to the right hand side is because t is fixed on the left hand side. However, on the right hand side we are taking the largest possible quantity as s varies in I.

The expression above implies that $\left|f_m(t) - f_n(t)\right| \leq \sup_{s \in I}\left|f_m(s) - f_n(s)\right| = \left\|f_m - f_n\right\|_\infty < \varepsilon$

This implies that the sequence $\{f_n(t)\}$ *is a Cauchy sequence in* $\mathbb{R}$.

$\{f_n(t)\}$ *must be convergent to some limit since it is a Cauchy sequence, and we know that Cauchy sequences of real numbers are convergent. The limit depends on t. Let's denote it by* $f(t)$. *Recall that* $f(t)$ *is a number. This shows that* f_n *converges to* f *pointwise.*

Now we must prove that f_n *converges to* f *also uniformly and not only pointwise.*

Moreover, we know that $\sup_{t \in I} \left| f(t) - f_n(t) \right| = \sup_{t \in I} \left[\lim_{m \to \infty} \left| f_m(t) - f_n(t) \right| \right]$,

because $f_m(t)$ converges to $f(t)$ as $m \to \infty$ since $f_n(t)$ converges pointwise to $f(t)$

We can manipulate the above expression to obtain:

$$\sup_{t \in I} \left| f(t) - f_n(t) \right| = \sup_{t \in I} \left[\lim_{m \to \infty} \left| f_m(t) - f_n(t) \right| \right] = \lim_{m \to \infty} \left(\sup_{t \in I} \left(\left| f_m(t) - f_n(t) \right| \right) \right)$$

Recall that $\forall\, m, n \geq N,\ \left| f_m(t) - f_n(t) \right| < \varepsilon$

This implies that $\sup_{t \in I} \left| f(t) - f_n(t) \right| = \lim_{m \to \infty} \left(\sup_{t \in I} \left(\left| f_m(t) - f_n(t) \right| \right) \right) \leq \varepsilon$, *which in turn implies that* $\sup_{t \in I} \left| f(t) - f_n(t) \right| \to 0$

To understand this last implication, consider a new sequence $\psi_n = \sup_{t \in I} \left| f(t) - f_n(t) \right|$. *For this sequence to converge to zero, it means that* $\forall \varepsilon > 0,\ \exists\ \widehat{N}$ *such that* $\forall n > \widehat{N},\ \left| \psi_n - 0 \right| = \left| \psi_n \right| < \varepsilon$

which is exactly what was shown above.

Another very important result is to show that the limit function $f(t)$ is also continuous. The proof of this result depends on uniform convergence and not on the sequence of functions being Cauchy.

To prove that $f(t)$ *is continuous, we have to show that* $\forall \varepsilon > 0,\ \exists\ \delta > 0$ *such that* $\left| t - a \right| < \delta$ *implies that* $\left| f(t) - f(a) \right| < \varepsilon$ *for every fixed* $a \in I$.

We begin by fixing $\varepsilon > 0$. *From before, since* $f_n(t)$ *converges to* $f(t)$ *uniformly as* $n \to \infty$, *this means that there exists an* N *such that* $\forall\, n \geq N$, $\left\| f_n - f \right\|_\infty < \dfrac{\varepsilon}{3}$.

Each $f_n(t)$ *is a continuous function because we started out with a family of continuous functions on the interval* I. *Fix now any* $n \geq N$, *this implies that* $\exists\, \delta$ *such that*

$$\left| t - a \right| < \delta \rightarrow \left| f_n(t) - f_n(a) \right| < \frac{\varepsilon}{3}$$

In addition, notice that $\left| f(t) - f(a) \right| \leq \left| f(t) - f_n(t) \right| + \left| f_n(t) - f_n(a) \right| + \left| f_n(a) - f(a) \right|$ (*Inequality* 1) *where we added and subtracted* $f_n(t)$ *and* $f_n(a)$ *and applied the triangle inequality*.

We can observe that $\left| f(t) - f_n(t) \right| \leq \left\| f - f_n \right\|_\infty$ *since* $\left\| f - f_n \right\|_\infty = \sup\limits_{t \in I} \left| f - f_n \right|$, *and by the same reason we can state that* $\left| f_n(a) - f(a) \right| < \left\| f - f_n \right\|_\infty$

After substituting all the elements highlighted in red above in place of inequality 1, *we obtain the following*:

$$\left| f(t) - f(a) \right| \leq \left| f(t) - f_n(t) \right| + \left| f_n(t) - f_{\left\| f - f_n \right\|_\infty n}(a) \right| + \left| f_n(a) - f(a) \right|$$

$$\left| f(t) - f(a) \right| \leq \left\| f - f_n \right\|_\infty + \frac{\varepsilon}{3} + \left\| f - f_n \right\|_\infty$$

$$\left| f(t) - f(a) \right| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$

$$\left| f(t) - f(a) \right| \leq \varepsilon$$

This concludes the proof of the theorem that states how a Cauchy set of continuous functions implies that they are uniformly convergent.

Chapter 5: Complete proof of the Picard-Lindelöf theorem

Up until this point, we gained an intuition behind differential equations, optimization, Cauchy sequences and Cauchy sequences of functions as well as the notions of Pointwise and Uniform convergence. The last step before moving on to the Euler-Lagrange differential equation is to understand the Picard-Lindelof theorem that provides specific conditions in which our Cauchy problem (initial value problem) has a unique solution in a specified domain. This theorem is useful because it has connections with many Calculus of Variations problems, for example the Brachistochrone problem. The Brachistochrone problem consists of a functional that we are trying to optimize as well as two initial boundary constraints; the Brachistochrone problem is an example of an initial value problem.

The optimization that we want to solve can be expressed in the following manner:

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases}$$

The problem that we want to solve above is known as a Cauchy problem.

Let's now consider a function f on a domain:

$$f : \left[t_0 - \tilde{a} \ , \ t_0 + \tilde{a}\right] \times \left[y_0 - b \ , \ y_0 + b\right] \to \mathbb{R}$$

$$\left[t_0 - \tilde{a} \ , \ t_0 + \tilde{a}\right] = I \quad and \left[y_0 - b \ , \ y_0 + b\right] \ = B$$

The function y solves the Cauchy problem, depicted above, if and only if it is differentiable and

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s)) \ ds \ , \ \forall \text{ t in some interval around } t_0$$

The reason is because if we plug t_0 into the expression above, the integral term equals zero and we are left with $y(t_0) = y_0$.

Let's define the function:

$\Gamma : \mathbb{C}(I,\ B)\ \rightarrow\ \mathbb{C}(I,\ B)$

$$\Gamma(\varphi)(t) = y_0 + \int_{t_0}^{t} f(s,\ \varphi(s))\ ds$$

The notation given by $\mathbb{C}(I,\ B)$, denotes that Gamma belongs to the set of continuous functions on the specified domain.
We want to look for a y $\in\ \mathbb{C}(I,\ \mathbb{R})$ *such that* $\Gamma(y) = y$. This is called a fixed point for the function Γ

The reason is because if we are able to find such a function y, this would solve the initial Cauchy problem.

Recall that a function f is Lipschitz in B, uniformly in t, if

$$\left|f\left(t,\ z_1\right) - f\left(t,\ z_2\right)\right| \leq\ L \cdot \left|z_1 - z_1\right|,\ \forall\, z_1,\ z_2 \in\ B \text{ and } \forall\, t \in\ I$$

Moreover, we now proceed in the following way:

Take $\varphi_1,\ \varphi_2 \in\ \mathbb{C}(I,\ \mathbb{R})$ *and consider*

$$\left\|\Gamma\left(\varphi_1\right) - \Gamma\left(\varphi_2\right)\right\|_{\infty} = \sup_{t \in\ I}\left|\int_{t_0}^{t}\left(f\left(s, \varphi_1(s)\right) - f\left(s, \varphi_2(s)\right)\right) ds\right|$$

$$\leq\ \sup_{t \in\ I}\int_{t_0}^{t}\left|f\left(s, \varphi_1(s)\right) - f\left(s, \varphi_2(s)\right)\right| ds$$

Since f *is Lipschitz, we know that*

$$\left\|\Gamma\left(\varphi_1\right) - \Gamma\left(\varphi_2\right)\right\|_{\infty}\ \leq\ \sup_{t \in I}\int_{t_0}^{t} L \cdot \left|\varphi_1(s) - \varphi_2(s)\right| ds$$

$$\leq\ \sup_{t \in I}\ \cdot L \cdot \sup_{s \in I}\left|\varphi_1(s) - \varphi_2(s)\right| \cdot \left(t - t_0\right)$$

The reason is because the

$$\sup_{s \in I} \left| \varphi_1(s) - \varphi_2(s) \right|$$

represents the largest possible value that the term $\varphi_1(s) - \varphi_2(s)$ can take. If we multiply it by $t - t_0$ (length of the interval) we are computing an upper bound for the area

$$\int_{t_0}^{t} \left| \varphi_1(s) - \varphi_2(s) \right| ds$$

Since $\left| t - t_0 \right| \leq \tilde{a}$, *we obtain that*

$$\left\| \Gamma(\varphi_1) - \Gamma(\varphi_2) \right\|_\infty \leq \tilde{a} \cdot L \cdot \left\| \varphi_1 - \varphi_2 \right\|_\infty \qquad \textbf{(Equation 5)}$$

At this point, fix $\xi_0 \in \mathbb{C}(I, \mathbb{R})$. *Define* $\xi_1 = \Gamma(\xi_0),\ \xi_2 = \Gamma(\xi_1),\ \dots\ ,\ \xi_n = \Gamma(\xi_{n-1})$

The goal is to build a sequence $\{\xi_n\}_{n \in \mathbb{N}}$ *of functions in* $\mathscr{C}(I, \mathbb{R})$

$$\begin{aligned} \left\| \xi_{n+1} - \xi_n \right\|_\infty &= \sup_{t \in I} \left| \Gamma(\xi_n)(t) - \Gamma(\xi_{n-1})(t) \right| \\ &= \left\| \Gamma(\xi_n) - \Gamma(\xi_{n-1}) \right\|_\infty \end{aligned}$$

From equation (1), *we know that*

$$\begin{aligned} \left\| \xi_{n+1} - \xi_n \right\|_\infty &\leq \tilde{a} \cdot L \cdot \left\| \xi_n - \xi_{n-1} \right\|_\infty \\ &= \tilde{a} \cdot L \cdot \left\| \Gamma(\xi_{n-1}) - \Gamma(\xi_{n-2}) \right\|_\infty \end{aligned}$$

By applying the same logic, *we know that*

$$\left\|\xi_{n+1}-\xi_n\right\|_\infty \le \tilde{a}\cdot L\cdot \tilde{a}\cdot L\cdot \left\|\xi_{n-1}-\xi_{n-2}\right\|_\infty$$

$$\le (\tilde{a}\cdot L)^2\cdot \left\|\xi_{n-1}-\xi_{n-2}\right\|_\infty$$

$$= (\tilde{a}\cdot L)^2\cdot \left\|\Gamma\left(\xi_{n-2}\right)-\Gamma\left(\xi_{n-3}\right)\right\|_\infty$$

If we continue applying the same procedure n times, we finally obtain:

$$\left\|\xi_{n+1}-\xi_n\right\|_\infty \le (\tilde{a}\cdot L)^n\cdot \left\|\xi_1-\xi_0\right\|_\infty \qquad \textbf{(Equation 5.1)}$$

Now, assume that $m > n$ *and consider*

$$\left\|\xi_m-\xi_n\right\|_\infty = \left\|\xi_m-\xi_{m-1}+\xi_{m-1}-\xi_{m-2}+\xi_{m-2}+\ \dots\ +\ \xi_{n+1}-\xi_n\right\|_\infty$$

$$\le \left\|\xi_m-\xi_{m-1}\right\|_\infty + \left\|\xi_{m-1}-\xi_{m-2}\right\|_\infty + \dots + \left\|\xi_{n+1}-\xi_n\right\|_\infty$$

Using what we figured out in (*Equation* 5.1), *we obtain*:

$$\left\|\xi_m-\xi_n\right\|_\infty \le \left[(\tilde{a}\cdot L)^{m-1}+(\tilde{a}\cdot L)^{m-2}+\dots+(\tilde{a}\cdot L)^n\right]\cdot\left\|\xi_1-\xi_0\right\|_\infty$$

The expression on the right hand side can be re − written in the following way:

$$\left[(\tilde{a}\cdot L)^{m-1}+(\tilde{a}\cdot L)^{m-2}+\dots+(\tilde{a}\cdot L)^n\right]\cdot\left\|\xi_1-\xi_0\right\|_\infty = (\tilde{a}\cdot L)^n\cdot\left\|\xi_1-\xi_0\right\|_\infty\cdot\sum_{j=0}^{m-n-1}(\tilde{a}\cdot L)^j$$

To better understand why, consider the following example:

Assume $m=5,\ n=2$

$$\left[(\tilde{a}\cdot L)^4+(\tilde{a}\cdot L)^3+(\tilde{a}\cdot L)^2\right] = (\tilde{a}\cdot L)^2\cdot\sum_{j=0}^{2}(\tilde{a}\cdot L)^j$$

Hence, we have figured out that:

$$\left\|\xi_m - \xi_n\right\|_\infty \leq (\tilde{a}\cdot L)^n \cdot \left\|\xi_1 - \xi_0\right\|_\infty \cdot \sum_{j=0}^{m-n-1} (\tilde{a}\cdot L)^j \leq (\tilde{a}\cdot L)^n \cdot \left\|\xi_1 - \xi_0\right\|_\infty \cdot \sum_{j=0}^{\infty} (\tilde{a}\cdot L)^j$$

because $\tilde{a}\cdot L > 0$

The series on the right hand side is in fact a geometric series. A geometric series $\sum_{j=0}^{\infty} r^j$ *converges to* $\frac{1}{1-r}$. *In our case, the series converges to* $\frac{1}{1-(\tilde{a}\cdot L)}$ *if* $|\tilde{a}\cdot L| < 1$, *that is if* $0 < \tilde{a}\cdot L < 1$

Hence, if $0 < \tilde{a}\cdot L < 1$ *then*

$$\left\|\xi_m - \xi_n\right\|_\infty \leq (\tilde{a}\cdot L)^n \cdot \left\|\xi_1 - \xi_0\right\|_\infty \cdot \frac{1}{1-(\tilde{a}\cdot L)}$$

Note $\tilde{a}\cdot L < 1$ *implied that* $\lim_{n\to\infty} (\tilde{a}\cdot L)^n = 0$. *More specifically,* $\forall\varepsilon$, $\exists N$ *such that*

$$\forall\, n \geq N,\ (\tilde{a}\cdot L)^n < \frac{(1-(\tilde{a}\cdot L))\cdot\varepsilon}{\left\|\xi_1 - \xi_0\right\|_\infty}.\ \textit{Hence, } \forall\, m,\, n \geq N\ \left\|\xi_m - \xi_n\right\|_\infty < \varepsilon \quad \textbf{(Equation 5.2)}$$

This implies that $\{\xi_n\}$ *is Cauchy which further implies that it is convergent.*

At this point, we denote by ξ *as the limit of the* ξ_n*'s, namely* $\lim_{n\to\infty} \xi_n = \xi$

Following the definition of ξ_n, *we can write* $\xi_n = \Gamma(\xi_{n-1})$. *Hence,* $\lim_{n\to\infty} \xi_n = \xi = \lim_{n\to\infty} \Gamma(\xi_{n-1})$.

Recall the following property:

If $H: \mathbb{R}\to\mathbb{R}$ *is a function and* $x_n \to x$, *then* $\lim_{n\to\infty} H(x_n) = H\left(\lim_{n\to\infty} x_n\right) = H(x)$ *if* H *is continuous*

We will now extend this definition concerning functionals.

DEFINITION:

An operator $A: \mathbb{C}(I, \mathbb{R}) \to \mathbb{C}(I, \mathbb{R})$ *is continuous if* $\forall\,\varepsilon > 0$, $\exists\,\delta > 0$ *such that*

$\|f - g\|_\infty < \delta$ *implies* $\|A(f) - A(g)\|_\infty < \varepsilon$, $\forall f, g$

The operator A in some ways varies the distance between f and g in a predictable manner.

$Lastly,\ if\ f_n \to f\ in\ \mathbb{C}(I,\ \mathbb{R})\ and\ A\ is\ continuous.\ Then,\ \forall \varepsilon > 0,\ \exists\ \delta\ such\ that$
$\left\| f_n - f \right\|_\infty < \delta\ implies\ \left\| A(f_n) - A(f) \right\|_\infty < \varepsilon.\ This\ means\ that\ \exists\ N\ such\ that\ \forall\ n \ge N,\ \left\| f_n - f \right\|_\infty < \delta$

Finally, consider **(equation 5)** given by

$$\left\| \Gamma(\varphi_1) - \Gamma(\varphi_2) \right\|_\infty \le \tilde{a} \cdot L \cdot \left\| \varphi_1 - \varphi_2 \right\|_\infty \qquad (\textbf{Equation 5})$$

$\forall \varepsilon > 0,\ we\ choose\ \delta < \dfrac{\varepsilon}{\tilde{a} \cdot L}$

We have proven that if f is Lipschitz, Γ has a fixed point that must be a solution to the initial value problem given by

$$\begin{cases} y'(t) = f(t,\ y(t)) \\ y(t_0) = y_0 \end{cases}$$

Recall that the function f at the beginning of this proof was defined in the following way:

$f : [t_0 - \tilde{a},\ t_0 + \tilde{a}] \times [y_0 - b,\ y_0 + b] \to \mathbb{R}$

$[t_0 - \tilde{a},\ t_0 + \tilde{a}] \in I\ \ and\ [y_0 - b,\ y_0 + b] \in B$

Knowing that f is Lipschitz, we automatically have a corresponding Lipschitz constant L. We figured out that for the geometric series to converge, we need $\tilde{a} \cdot L < 1$. To have this condition satisfied we may need to restrict the domain of f. We restrict I to $[t_0 - a, t_0 + a]$ so that $a < \dfrac{1}{L}$

Since we want $\Gamma(\varphi) \in \mathbb{C}(I, B)$, w e need that

$$\Gamma(\varphi)(t) \in B,\ \forall\ t$$

which is equivalent to

$$\left|\Gamma(\varphi)(t) - y_0\right| \leq b$$

Moreover, f being Lipschitz implies that f is continuous. The $Domain(f) = I \times B$. Since the Domain of f is a closed and bounded set (it is a rectangle), we know by the Weierstrass theorem that f has a maximum on $I \times B$. Let's denote this maximum by

$$M = \max_{I \times B} f$$

$$\text{We have } \left|\int_{t_0}^{t} f(s, \varphi(s))\, ds\right| \leq \int_{t_0}^{t} \left|f(s, \varphi(s))\right| ds \leq M \cdot \left(t - t_0\right) \leq b$$

$\text{Therefore, we need to choose } a \text{ such that } a \leq \dfrac{b}{M} \text{ for the the inequality to hold.}$

Putting everything together, we need that

$$a < \min\left\{\frac{1}{L}, \frac{b}{m}, \tilde{a}\right\}$$

This result implies that we can solve

$y'(t) = f(t, y(t)),\ y\left(t_0\right) = t$ *on the modified interval* $\left[t_0 - a, t_0 + a\right]$ *which is usually smaller than* $\left[t_0 - \tilde{a}, t_0 + \tilde{a}\right]$.

PART 2

Chapter 6: Optimization intuition

In this chapter of the thesis, we will explore the general idea of the concept of optimization which is a central pillar to understand the idea of Calculus of Variations. When the terms "minimize" or "maximize" are used, this is a clear indication that we are dealing with some sort of optimization problem. The idea of optimizing a function can be generalized to any multivariable function, and not solely to single variable functions. Partial derivatives are employed when trying to optimize multivariable functions analogously to how optimization is conducted for single variable functions.

We will briefly explore the concept of unconstrained optimization for functions of several variables and explain some important results and definitions. Next, we will move on to Constrained Optimization dealing with First Order Conditions. Lastly, we will analyze the Lagrange Multiplier method applied to optimization problems.

The definition of a maximum and minimum for functions of several variables is analogous to functions of a single variable. Before moving on to these definitions, let's explore the meaning of the $B_r(\hat{x})$ notation.

***DEFINITION* ($B_r(\tilde{x})$ *notation*)**

$B_r(\tilde{x}) = \{x \in U \mid d(x, \tilde{x}) < r\}$ *where* $r > 0$

The above definition denotes a ball centered at $\tilde{x}$ *having radius* $r > 0$

***DEFINITION* (Maximum and Minimum of multivariable functions)**

$F: U \rightarrow \mathbb{R}$ *is a function of* n *variables that has a domain* $U \subseteq \mathbb{R}^n$:

(1) *An element* $\hat{x} \in U$ *is a maximum of* F *if* $F(\hat{x}) \geq F(x), \forall x \in U$

(2) *An element* $\hat{x} \in U$ *is a strict maximum of* F *if* $F(\hat{x}) > F(x), \forall x \in U$ *such that* $\hat{x} \neq x$

(3) *An element* $\hat{x} \in U$ *is a local maximum of* F *if there exists a ball* $B_r(\hat{x})$ *encompassing* $\hat{x}$, *such that*

$F(\hat{x}) \geq F(x), \forall x \in B_r(\hat{x}) \cap U$.

(4) *An element* $\hat{x} \in U$ *is a strict local maximum of* F *if there exists a ball* $B_r(\hat{x})$ *encompassing* $\hat{x}$, *such that* $F(\hat{x}) > F(x), \forall x \in B_r(\hat{x}) \cap U$ *such that* $x \neq \hat{x}$

For a point to be a local and/or strict minimum, the same definition given above applies, with the inequality sign reversed.

What condition three is stating is that a point is a local max for a function if there are no nearby points in which the function takes on a larger value. It is important to clarify that the words "nearby points" denote a neighborhood around the candidate point for being a local max. The neighborhood is typically chosen to be a ball having radius r. Mathematically, a neighborhood is defined in the following way:

***DEFINITION* (neighborhood)**

A neighborhood of a point $\tilde{x}$ *is any set* V *such that* $\exists\ r > 0$ *in which* $B_r(\tilde{x}) \in V$

Additionally, let's gain some intuition regarding what a saddle point is.

***DEFINITION* (Saddle point)**

If $f: \mathbb{R}^n \to \mathbb{R}$ *where* $n \geq 2$, *a point* $\tilde{x} \in \mathbb{R}^n$ *is a saddle point if* $\forall$ *neighborhood of point* $\tilde{x}$ *with* $r > 0$

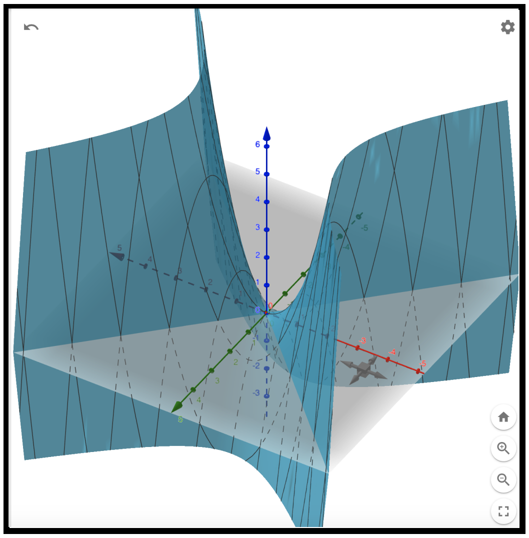
there exist points

(1) x_0 *belonging to* $\mathbb{R}^n$ *as well as to the neighborhood of* x_0 *such that* $f\left(x_0\right) \leq f(\tilde{x})$

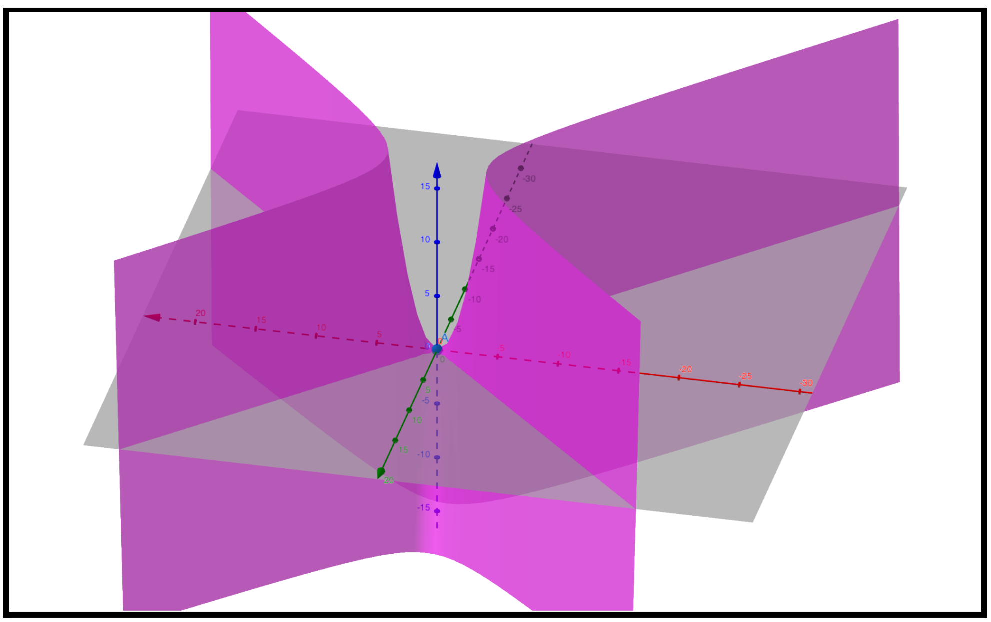
(2) x' *belonging to* $\mathbb{R}^n$ *as well as to the neighborhood of* x' *such that* $f(x') \geq f(\tilde{x})$

A saddle point is a critical point in which, along a certain direction, the function attains a minimum, while along another direction, the function attains a maximum. Let's look at some visuals:

The function $f(x, y) = x^2 - y^2$ has a saddle point at $(0, 0)$. **Figure 4 provides a visual representation.**



(Figure 4)



(Figure 5)

Notice that the function at $(0, 0)$ has a saddle point. In fact, the second order partial derivatives evaluated at that point have opposite signs, and this phenomenon will be discussed in detail later on in the thesis.

Moreover, when discussing Unconstrained Optimization for multivariable functions, it is critical to generalize the First Order and Second Order conditions, applied to single variable functions, to multivariable ones. Before moving to the theorem, let's briefly look at the definition of what an interior point is, since this concept will be incorporated in the theorem.

DEFINITION **(interior point)**

A point $\tilde{x}$ *is an interior point of set* U, *if* $\exists\ r > 0$ *such that* $B_r(\tilde{x}) \subset U$

THEOREM

$F: U \rightarrow \mathbb{R}^1$ *is a* C^1 *function that has as its domain a set* $U \subseteq \mathbb{R}^n$.

Then, if $\hat{x}$ *is is a local maximum or minimum for* F *on the domain* U, *and* $\hat{x}$ *is an interior point of* U, *this implies that:*

$$\frac{\partial F}{\partial x_i}(\hat{x}) = 0\ ,\ \forall\ i = 1,\ 2,\ \ldots\ ,\ n$$

Now that we have explored the First Order Condition for functions of n variables, we can explore the Second Order Conditions. But, before moving on, it is crucial to describe a very important theorem called the Schwarz Theorem. This theorem provides a condition that guarantees that the Hessian matrix is symmetric.

THEOREM **(Schwarz mixed partial derivatives theorem)**

A function $F: \theta \rightarrow \mathbb{R}$ *is defined on a set* $\theta \subset \mathbb{R}^n$. *If* x *is an interior point of* θ

and F *has continuous second order partial derivatives at* x, *then*

$\forall\ i,j \in \{1, 2, 3,\ \ldots\ ,\ n\}\ ,\ F_{x_i x_j}(x) = F_{x_j x_i}(x)$

Let’s now explore the Second Order Conditions.

Given a function of n variables, the Hessian matrix is a symmetric matrix that encapsulates all the second order partial derivatives of the function. More specifically, the Hessian matrix evaluated at a generic point x is defined as:

$$H(F(x)) = \begin{bmatrix} \frac{\partial^2 F(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 F(x)}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 F(x)}{\partial x_n^2} \end{bmatrix}$$

***DEFINITION* ($\mathbb{C}^1$ functions)**

$\mathbb{C}^1$ *denotes the class of all differentiable functions whose derivative is also continuous*

***DEFINITION* ($\mathbb{C}^2$ functions)**

$\mathbb{C}^2$ *denotes the class of all functions that are twice continuous and differentiable*

Let's consider an example:

EXAMPLE 7 (Hessian Matrix)

Consider the function $F:\mathbb{R}^2 \to \mathbb{R}$, such that $F(x, y) = x^2 + yx^3$

Compute the Hessian Matrix of $F(x, y)$.

From the Schwarz theorem, since function $F(x, y)$ is a polynomial, we know that its second order partial derivatives are continuous, hence the Hessian matrix will be symmetric.

We begin by computing the first order, second order and mixed partial derivatives of the function:

$$\frac{\partial F}{\partial x} = 2x + 3yx^2 \qquad \frac{\partial^2 F}{\partial x^2} = 2 + 6xy$$

$$\frac{\partial F}{\partial y} = x^3 \qquad \frac{\partial^2 F}{\partial y^2} = 0$$

$$\frac{\partial^2 F}{\partial x \partial y} = 3x^2$$

The Hessian is given by:

$$H(F(x)) = \begin{bmatrix} 2 + 6xy & 3x^2 \\ 3x^2 & 0 \end{bmatrix}$$

The Hessian matrix is significant because by classifying the Hessian matrix of a multivariable function evaluated at a critical point, we can understand the nature of this critical point, that is if the point corresponds to a local max, min or a saddle point.

We will always assume that the Hessian matrix is symmetric.

THEOREM (sufficient conditon)

Let $F: U \to \mathbb{R}^1$ *be a* $\mathbb{C}^2$ *function. Its domain* $U \subseteq \mathbb{R}^n$. *Moreover,* $\hat{x}$ *is a critical point which implies that* $\frac{\partial F}{\partial x_i}(\hat{x}) = 0$, $\forall\, i = 1, 2, \ldots, n$.

We can make the following statements:

(1) *If* $H(F(\hat{x}))$ *is negative definite, then* $\hat{x}$ *is a strict local max.*

(2) *If* $H(F(\hat{x}))$ *is positive definite, then* $\hat{x}$ *is a strict local min.*

(3) *If* $H(F(\hat{x}))$ *is indefinite, then* $\hat{x}$ *is neither a local max nor a local min.* $\hat{x}$ *is a saddle point.*

The theorem that we stated above concerning the sufficient conditions for a critical point to be a local maximum, minimum or saddle point can be reformulated in another more convenient way, using properties of symmetric matrices. This new formulation of the theorem is based on Theorem 17.3 from the book “Mathematics for Economists" by Carl P.Simon and Lawrence E.Blume. Before moving on to the theorem, let’s explore two important definitions.

***DEFINITION* (Principal Minor)**

A principal minor of order k is the submatrix that results after deleting any $n-k$ rows and their corresponding columns from an $n \times n$ matrix A.

***DEFINITION* (Leading Principal Minor)**

A leading principal minor of order k of an $n \times n$ matrix A results after deleting the last $n-k$ rows and their corresponding columns.

Now, we are ready to explore the theorem below.

THEOREM (Sufficient condition reformulated)

Let $F: U \to \mathbb{R}^1$ be a C^2 function that has a domain $U \subseteq \mathbb{R}^n$ which is an open set.

Given that $\frac{\partial F}{\partial x_i}(\hat{x}) = 0$, $\forall\, i = 1, 2, \ldots, n$ and if all the leading principal minors of $H(F(\hat{x}))$ alternate in sign $(-, +, -)$, then $\hat{x}$ is a strict local max.

On the other hand, if all the leading principal minors of $H(\hat{x})$ are positive, this implies that $\hat{x}$ is a strict local minimum.

If there are some non-zero leading principal minors of $H(\hat{x})$ that violate the sign patterns of before, this implies that $\hat{x}$ is a saddle point.

Like for functions of single variables, we can generalize the second order necessary conditions for multivariable functions. Below, we will explore the necessary conditions and then we will provide a concrete example.

THEOREM (necessary conditions)

$F: U \rightarrow \mathbb{R}^1$ is a $\mathbb{C}^2$ function having n variables as input. Moreover, $\hat{x}$ is an interior point of U, and $U \subseteq \mathbb{R}^n$.

(1) If $\hat{x}$ is a local minimum of F, then this implies that $\frac{\partial F(\hat{x})}{\partial x_i} = 0$, $\forall i = 1, \ldots, n$ and all the principal minors of the $H(F(\hat{x}))$ are ≥ 0.

(2) If $\hat{x}$ is a local maximum of F, then this implies that $\frac{\partial F(\hat{x})}{\partial x_i} = 0$, $\forall i = 1, \ldots, n$ and all the principal minors of the $H(F(\hat{x}))$ of odd order are are ≤ 0 and all principal minors of even order are ≥ 0.

Let's look at a concrete example with some visuals:

EXAMPLE 8 (Unconstrained optimization)

Consider the function $F(x, y) = x^3 - y^3 + 9xy$

Find and classify its critical points

We begin by computing its first order and second order partial derivatives:

$$\frac{\partial F(x, y)}{\partial x} = 3x^2 + 9y$$

$$\frac{\partial F(x, y)}{\partial y} = -3y^2 + 9x$$

We then set both partial derivatives equal to zero. Since the partial derivatives are continuous because $F(x, y)$ is a polynomial, this means the function $F(x, y)$ is differentiable, and therefore, the tangent plane is well defined at the points in which the partial derivatives are equal to zero.

$$3x^2 + 9y = 0 \quad (\textbf{equation 6})$$

$$-3y^2 + 9x = 0 \quad (\textbf{equation 6.1})$$

We then try to solve equations 6 and 6.1:

$$9y = -3x^2 \rightarrow y = -\frac{1}{3}x^2 \quad \text{by manipulating } (\textbf{equation 6})$$

$$9x = 3y^2 \rightarrow 3x = y^2 \quad \text{by manipulating } (\textbf{equation 6.1})$$

After we substitute **(equation 6)** into **(equation 6.1)** for the variable y, we obtain:

$$3x = \frac{1}{9}x^4$$

$$27x - x^4 = 0$$

$$x \cdot \left(27 - x^3\right) = 0$$

The two possible solutions are: $x = 0$, $x = 3$

To find the corresponding y coordinates, we substitute both values of x that we found above into equation 6:

If $x = 0,\ y = -\frac{1}{3} \cdot (0)^2 = 0$

If $x = 3,\ y = -\frac{1}{3} \cdot (9) = -3$

The candidate points are $(0, 0)$ and $(3, -3)$

We must now compute the Hessian matrix evaluated at both candidate points in order to classify both candidate points:

$$\frac{\partial^2 F(x, y)}{\partial x^2} = 6x$$

$$\frac{\partial^2 F(x, y)}{\partial y^2} = -6y$$

$$\frac{\partial^2 F(x, y)}{\partial y \partial x} = 9$$

We can now construct the Hessian matrix:

$$H(F(x)) = \begin{bmatrix} 6x & 9 \\ 9 & -6y \end{bmatrix}$$

Let's evaluate the Hessian at the critical points that we computed above:

$$H(F(0, 0)) = \begin{bmatrix} 0 & 9 \\ 9 & 0 \end{bmatrix}$$

Let's apply the theorem containing the sufficient conditions. We have to compute the leading principal minors which are different from the principal minors:

$$Leading\ principal\ minor_1 = 0$$

$$Leading\ principal\ minor_2 = 0 - 9 \cdot 9 = -81$$

Since the second leading principal minor violates the sign pattern $(-, +, -, \ldots)$, by the sufficient condition theorem, $(0, 0)$ is a saddle point.

Moreover, let's do the same for the other point:

$$H(F(3, -3)) = \begin{bmatrix} 18 & 9 \\ 9 & 18 \end{bmatrix}$$

$$Leading\ principal\ minor_1 = 18$$

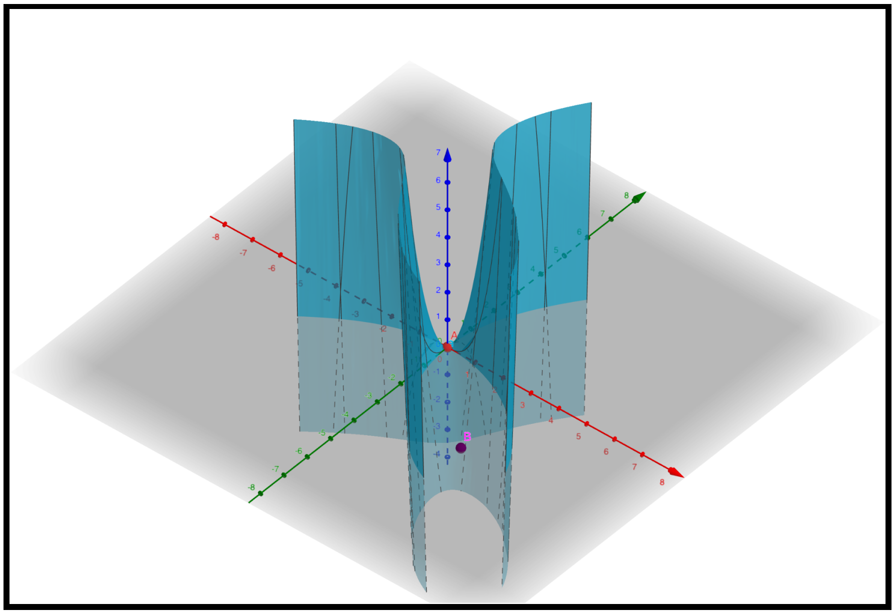
$$Leading\ principal\ minor_2 = 18^2 - 81 = 243$$

By the sufficient condition theorem, since all leading principal minors are positive, this means that $(3, -3)$ is a local minimum.

Moreover, $(3, -3)$ is not a global minimum because of the following observation:

If we pick a path $x = 0$ along $F(x, y)$ and take the limit as $y \to \infty$, we would get:

$$\lim_{y \to \infty} F(0, y) = -y^3 = -\infty$$



(Figure 6)

Figure 6 provides a visual representation of the function. Points A and B are plotted which correspond to a saddle point and local minimum respectively.

So far we have explored the concept of unconstrained optimization for functions of several variables. The last part of this section of the thesis, concerning optimization, will explore the concept of Lagrange Multipliers applied to a Constrained optimization problem subject to one or more equality constraints.

Before moving on to the concept of Lagrange Multipliers, we will briefly define what a level curve for a function is.

***DEFINITION* (level curve)**

Let $f : \mathbb{R}^n \to \mathbb{R}$. The level curves of f are defined as the curves having equation

$$f\left(x_1, x_2, \ldots, x_n\right) = k \ \text{ where } k \in \mathbb{R}.$$

Now, consider the following problem:

Maximize $f\left(x_1, x_2\right)$ subject to $h\left(x_1, x_2\right) = D$, where $C_1, C_2, D \in \mathbb{R}$

The general idea is that we want to find a specific point x^* such that the level curve of f at

$$\mathbf{x}^* = \left(x^*, y^* \right) \in Domain(f)$$

is tangent to the curve given by $C_1x_1 + C_2x_2 = D$. Thanks to this tangency condition at $\boldsymbol{x}^*$, the gradient of the function $f(x_1, x_2)$ is equal to the gradient of the constraint function $C_1x_1 + C_2x_2 = D$. More specifically, assuming that

$$\frac{\partial f\left(\mathbf{x}^* \right)}{\partial x_2}, \frac{\partial h\left(\mathbf{x}^* \right)}{\partial x_2} \neq 0$$

we can write the following:

$$\frac{\frac{\partial f\left(\mathbf{x}^* \right)}{\partial x_1}}{\frac{\partial f\left(\mathbf{x}^* \right)}{\partial x_2}} = \frac{\frac{\partial h\left(\mathbf{x}^* \right)}{\partial x_1}}{\frac{\partial h\left(\mathbf{x}^* \right)}{\partial x_2}}$$

The statement above is equivalent to the following equation:

$$\frac{\partial f\left(\boldsymbol{x}^* \right)}{\partial x_1} - \mu \cdot \frac{\partial h\left(\boldsymbol{x}^* \right)}{\partial x_1} = 0$$

$$\frac{\partial f\left(\boldsymbol{x}^* \right)}{\partial x_2} - \mu \cdot \frac{\partial h\left(\boldsymbol{x}^* \right)}{\partial x_2} = 0$$

The parameter μ is called the Lagrange Multiplier.

Since we are trying to solve for three unknown quantities, namely x_1, x_2, μ we need another equation. Such equation is given by the constraint itself:

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_1} - \mu \cdot \frac{\partial h(\mathbf{x}^*)}{\partial x_1} = 0$$

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_2} - \mu \cdot \frac{\partial h(\mathbf{x}^*)}{\partial x_2} = 0$$

$$h(x_1, x_2) - D = 0$$

The three equations above can be encapsulated into the so-called Lagrangian equation. We define a Lagrangian function

$$L: Domain(f) \times \mathbb{R} \to \mathbb{R}$$

in the following way:

$$L(x_1, x_2, \mu) = f(x_1, x_2) - \mu \cdot \left[h(x_1, x_2) - D\right]$$

and we solve $\nabla L = 0$

Solving the initial constrained optimization problem with an equality constraint is equivalent to maximizing the Lagrangian function by setting each one of its partial derivatives equal to zero, as if it was an unconstrained optimization problem that we covered in the previous section.

In the above discussion, it is crucial that at least one of the two partial derivatives

$$\frac{\partial h(\mathbf{x}^*)}{\partial x_1}, \frac{\partial h(\mathbf{x}^*)}{\partial x_2}$$

are different from zero. The condition is called a **Constraint qualification**. The reason is because at a critical point, we want the constraint function and the level curve of f to be tangent to each other. If both partial derivatives of the constraint function at that point are zero, then this implies that the constraint function is not a curve.

Let’s now consider a concrete example in which we apply the Lagrange Multipliers Theorem:

EXAMPLE 10 (Lagrange Multipliers – constrained optimization)

Maximize $f(x_1, x_2) = x_1^2 \cdot x_2 + x_1 \cdot x_2 + x_2^2$ subject to $h(x_1, x_2) = x_1^2 + x_2^2 = 8$

We can solve this constrained optimization problem by applying the Lagrange Multiplier theorem;

$$\nabla f = \mu \cdot \nabla h$$

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \mu \cdot \frac{\partial h(x_1, x_2)}{\partial x_1}$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = \mu \cdot \frac{\partial h(x_1, x_2)}{\partial x_2}$$

$$2 \cdot x_1 + x_2 = \lambda \cdot 2 \cdot x_1$$

$$x_1 + 2 \cdot x_2 = \mu \cdot 2 \cdot x_2$$

We will try to eliminate the parameter μ:

$$\frac{2 \cdot x_1 + x_2}{2 \cdot x_1} = \frac{x_1 + 2 \cdot x_2}{2 \cdot x_2}$$

$$2 \cdot x_1 \cdot x_2 + x_2^2 = x_1^2 + 2 \cdot x_1 \cdot x_2$$

$$x_2^2 = x_1^2$$

$$x_2 = \pm\ x_1 \quad \textbf{(equation 6.2)}$$

Moreover, substituting **(equation 6.2)** into the original constraint function, we obtain the following:

$$x_1^2 + x_1^2 = 8$$

$$2 \cdot x_1^2 = 8$$

$$x_1^2 = 4 \rightarrow x_1 = \pm\ 2$$

This implies that the candidate points are:

$(2, 2) \quad (2, -2) \quad (-2, -2) \quad (-2, 2)$

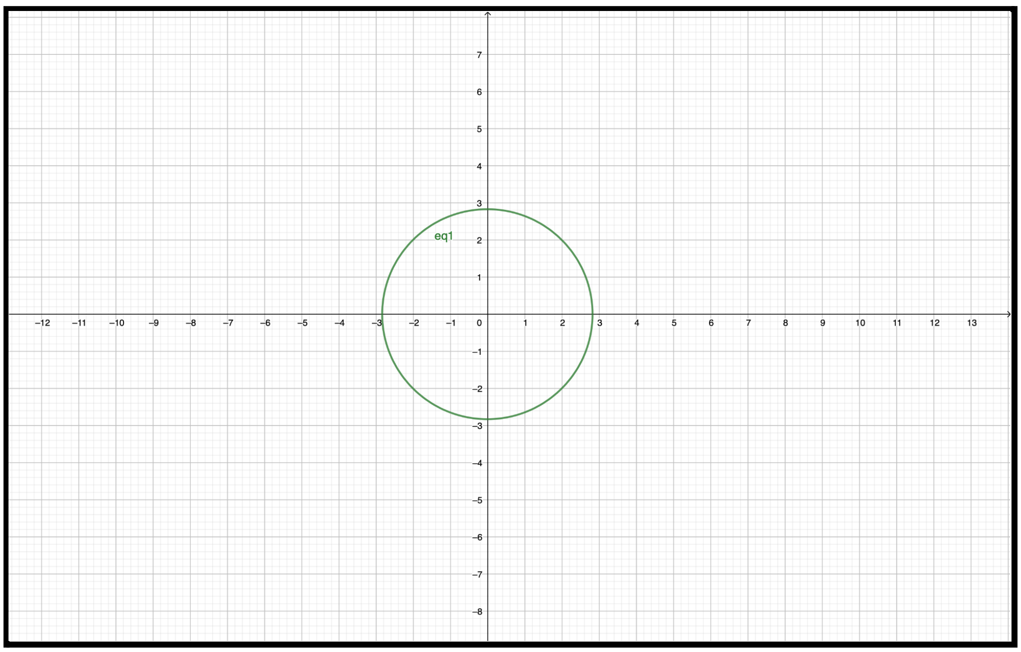
Substituting these points into the function $f(x_1, x_2)$ we obtain:

$(2, 2, 12) \quad (2, -2, 4) \quad (-2, 2, 4) \quad (-2, -2, 12)$

The points that maximize the function subject to the constraint are:

$$(2, 2, 12) \text{ and } (-2, -2, 12)$$

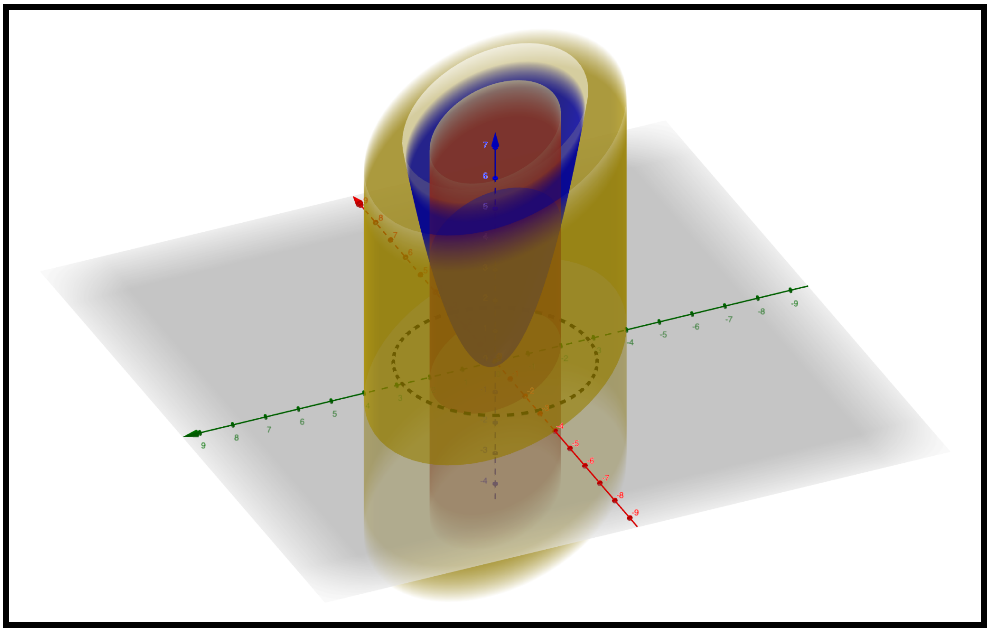
Now, let's gain some intuition regarding what we just computed.



(Figure 7)

Figure 7 provides a graph of the constraint function.

We are essentially asking, out of all possible level curves of the function $f(x_1, x_2)$, which level curve is tangent to the constraint function depicted in **Figure 7.**

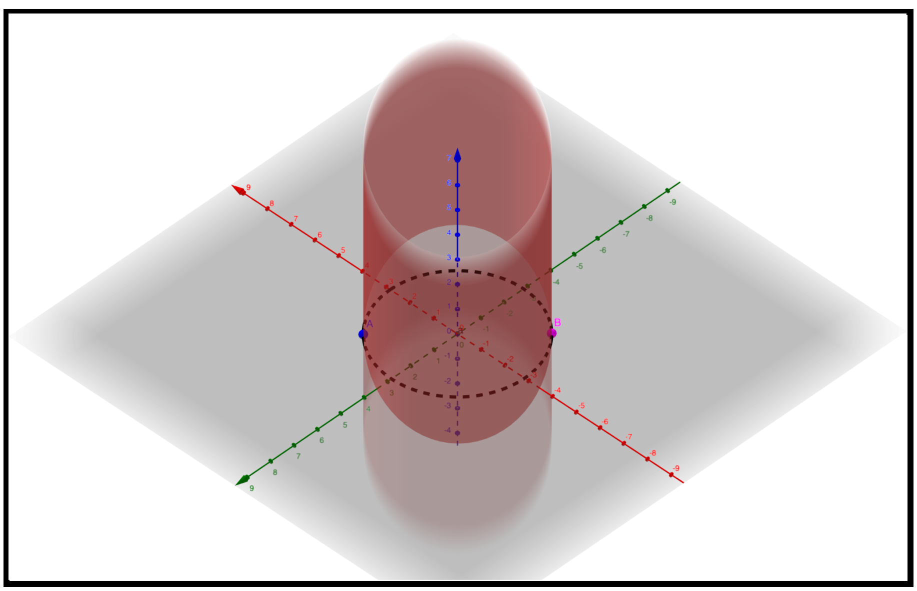


(Figure 8)

Figure 8 depicts several level curves of the function $f(x_1, x_2)$ drawn in yellow and red. Moreover, the constraint function is drawn in black (black dotted lines) and the function $f(x_1, x_2)$ is depicted in blue.

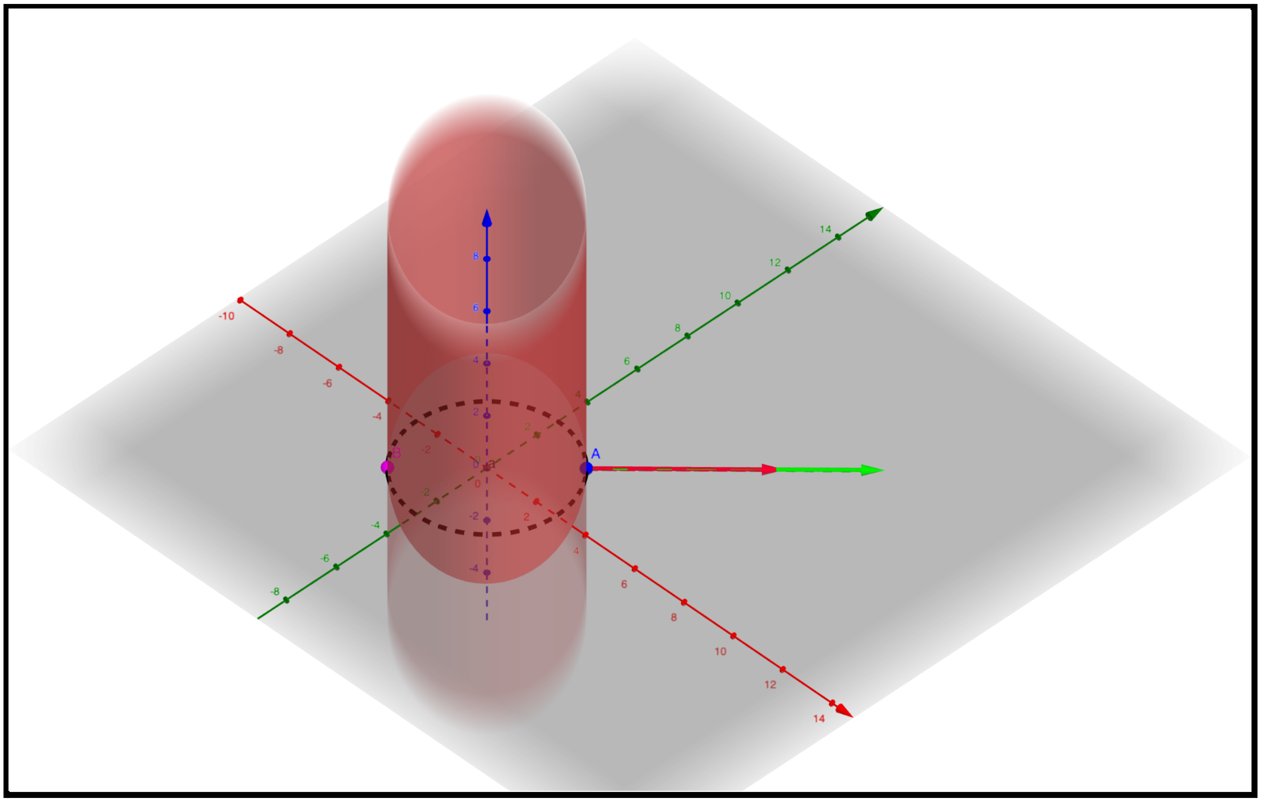
We found that the level curve associated to the points $(2, 2)$, $(-2, -2)$ is given by:

$$x_1^2 + x_1 \cdot x_2 + x_2^2 = 12$$



(Figure 9)

In **Figure 9** we have drawn in red the level curve associated to the points $(2, 2)$, $(-2, -2)$ and in black, the constraint function. We also plotted the points $(2, 2)$, $(-2, -2)$ denoted by A and B respectively.



(Figure 10)

In **Figure 10,** It is evident that the gradient vector of the constraint function at the point $A\ (2, 2)$, drawn in red, is a scalar multiple of the gradient vector of the level curve of $f(x_1, x_2)$ at the point $A\ (2, 2)$, drawn in green.

In some cases, we can deal with constrained optimization problems with more than one equality constraint. More specifically:

Given a function $f\left(x_1, x_2, \ldots, x_n\right)$ we want to maximize it subject to a constraint set defined by $\theta_h = \left\{ \mathbf{x} \in \left(x_1, \ldots, x_n\right) \middle| h_1(\mathbf{x}) = a_1, h_2(\mathbf{x}) = a_2, \ldots, h_m(\mathbf{x}) = a_m \right\}$

We must also generalize the constraint qualification involving constrained optimization subject to one equality constraint.
In the case that we are dealing with just one equality constraint, the constraint qualification is met if there are some first order partial derivatives of the constraint function evaluated at the critical point that are different from zero. In this case, we have to generalize this condition using the Jacobian matrix. First, we will briefly define the Jacobian matrix.

***DEFINITION* (Jacobian Matrix)**

If we consider a function $F: \mathbb{R}^n \to \mathbb{R}^m$ *such that* $n, m \in \mathbb{N}$ *and if each of the function's first order partial derivatives exist in a neighborhood of* $x \in \mathbb{R}^n$, *then the Jacobian of* F *is defined as the matrix* $J_{m \times n}(F(x))$

whose (i,j) *entry is* $J_{ij} = \dfrac{\partial F_i}{\partial x_j}(x)$

$$\textit{More specifically, } J_{m\times n}(F(x)) = \begin{bmatrix} \dfrac{\partial F_1}{\partial x_1}(x) & \dots & \dots & \dfrac{\partial F_n}{\partial x_n}(x) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1}(x) & \dots & \dots & \dfrac{\partial F_m}{\partial x_n}(x) \end{bmatrix}$$

It is crucial to notice that F *takes a point* $\boldsymbol{x} \in \mathbb{R}^n$ *and outputs a vector* $F(\boldsymbol{x}) \in \mathbb{R}^m$.

Now, we can resume our discussion above, since we have defined the Jacobian matrix.

If the constraints are $m > 1$, *we need to introduce the Jacobian matrix*:

$$J(h(\boldsymbol{x}^*)) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(\boldsymbol{x}^*) & \cdots & \cdots & \frac{\partial h_1}{\partial x_n}(\boldsymbol{x}^*) \\ \frac{\partial h_2}{\partial x_2}(\boldsymbol{x}^*) & \cdots & \cdots & \frac{\partial h_2}{\partial x_n}(\boldsymbol{x}^*) \\ \vdots & \cdots & \cdots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\boldsymbol{x}^*) & \cdots & \cdots & \frac{\partial h_m}{\partial x_n}(\boldsymbol{x}^*) \end{bmatrix}$$

The constraint functions $\left(\mathrm{h}_1, \ldots, \mathrm{h}_\mathrm{m}\right)$ satisfy the nondegenerate constraint qualification at $\mathbf{x}^*$ if the rank of $\mathrm{J}\left(\mathrm{h}\left(\mathbf{x}^*\right)\right)$ is m .

We are ready to explore the theorem that explains how to apply the techniques of Lagrange Multipliers to a constrained optimization problem with more than one constraint function:

THEOREM *(Lagrange Multipliers with m equality constraints)*

f *and* $h_1, h_2, \ldots, h_m$ *are* C^1 *functions taking on n variables as input.*

Moreover, we want to maximixe $f(\boldsymbol{x})$ *on the set*

$$\theta_h = \left\{ x \in \left(x_1, \ldots, x_n\right) \middle| h_1(\boldsymbol{x}) = a_1, h_2(\boldsymbol{x}) = a_2, \ldots, h_m(\boldsymbol{x}) = a_m \right\}$$

Additionally, we know that $\boldsymbol{x}^* \in \theta_h$ *and that* $\boldsymbol{x}^*$ *is a local maximum or minimum of* f *subject to* θ_h. $\boldsymbol{x}^*$ *also satisfies the nondegenerate constraint qualification that we explained above. This implies that* $\exists\ \mu_1^*, \ldots, \mu_m^*$ *such that* $\left(x_1^*, \ldots, x_n^*, \mu_1^*, \ldots, \mu_m^*\right) = \left(\boldsymbol{x}^*, \boldsymbol{\mu}^*\right)$ *is a critical point of the Lagrangian function.*

Recall that $L(\boldsymbol{x}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \mu_1 \cdot \left[h_1(\boldsymbol{x}) - a_1\right] - \ldots - \mu_m \cdot \left[h_m(\boldsymbol{x}) - a_m\right]$

Chapter 7: First and Second Variation of a functional

In this chapter of the thesis, we will explore the concept of a functional. Next, we will derive the First Variation as well as the Second Variation of a functional using an interesting approach based on a Taylor Series approximation. Lastly, we will explore Legendre's necessary condition for a weak minimum based on the Second Variation of a functional.

In the previous sections of this thesis, we gained an intuition behind the concept of optimization involving a real-valued function in $\mathbb{R}^n$, and we tried to find points that maximized or minimized the function subject to some constraints. Now, we will move towards a new idea, that is trying to find a function belonging to a space of functions, that minimizes a functional. More specifically, V denotes a vector space containing functions. $y \in V$ is an arbitrary function in V.

We call a functional J, a real valued function defined on V, that is

$$J: V \rightarrow \mathbb{R}$$

When dealing with optimization of real-valued functions, we know that we can approximate a multivariable function by generalizing the concept of a Taylor expansion to $\mathbb{R}^n$. We essentially pick a point, and then we try to approximate the function locally evaluated at that point. We essentially take increments and add them to the function evaluated at that initial point in a neighborhood around that point. To take increments we need a distance.

In this case, how do we define the concept of distance for functions ? We will first define the concept of a norm because the norm gives us a distance which allows us to define local maxima and minima, and enables us to analyze concepts like convergence and continuity.

***DEFINITION* (Norm on the space V)**

A norm on V is a function $\left\|\cdot\right\|: V \rightarrow \mathbb{R}$, such that

(1) $\|y\| > 0$ *if* $y \neq 0$

(2) $\|y\| = 0$ *if* $y = 0$.

(3) $\|\lambda \cdot y\| = |\lambda| \cdot \|y\|, \ \forall \lambda \in \mathbb{R}, \ \forall y \in V$

(4) $\|y + z\| \leq \|y\| + \|z\| , \ \forall y, z \in V$

In other words, the norm must obey the four properties described above. The definition of the norm given above makes sense only if V is a vector space. Given a norm $\|\cdot\|$ on V we can define the distance between any two elements $y, z \in V$ as

$$\|y - z\|$$

We can now define what it means for a function $y \in V$ to minimize a functional J defined on a vector space V. We need to have a clear definition of the concept of distance, which we have given above, to define concepts like local maxima and minima of functionals. With a norm, hence a distance, we can also define balls around points:

$$For\ r > 0\ and\ y \in V,\ we\ define\ B_r(y) = \{z \in V \mid \|z - y\| < r\}$$

DEFINITION

$V\ is\ a\ vector\ space\ of\ funcions\ having\ norm\ \|\cdot\|.$

$Moreover,\ J\ is\ a\ functional\ on\ V.\ A\ function\ y^* \in V\ is\ a\ local\ min\ for\ J\ if\ \exists\ \varepsilon > 0$

$such\ that\ \forall\, y \in B_\varepsilon(y^*)\ ,we\ have\ that\ J(y^*) \le J(y)$

What the definition above is saying is that if we consider a ball of radius ϵ, and analyze all functions y in that neighborhood, $J(y^\star)$ is smaller than or equal to $J(y)$.

We can now move on to deriving the First Variation of a functional.

The method that will be used has a direct parallelism with the derivation of the first Order necessary condition concerning unconstrained optimization of a real valued function on $\mathbb{R}^n$. We propose here a particular proof of the Sufficient Condition theorem given on page 43 to showcase how this same logical reasoning can be applied to obtain the First Variation of a functional.

We begin by fixing a domain $D \subset \mathbb{R}^n$. Moreover, we also assume that:

$$f \in \mathbb{C}^1\ and\ x^*\ is\ a\ local\ min\ for\ the\ function\ f$$

We consider $d \in \mathbb{R}^n$ being an arbitrary vector, and we assume that

$$x^* + \alpha \cdot d \in D,\ \forall\, \alpha \in \mathbb{R}\ that\ are\ close\ to\ 0$$

We are essentially starting at the point x^*, and then we are moving along the vector d with increments equal to α.

When d is fixed, we define the following:

$$g(\alpha) = f\left(x^* + \alpha \cdot d\right)$$

Notice that g is differentiable because f is differentiable. We now begin by constructing the Taylor expansion centered around $\alpha = 0$ for the function $g(\alpha)$:

$$g(\alpha) = g(0) + g'(0) \cdot \alpha + \psi(\alpha) \qquad \textbf{(Equation 7)}$$

The term $\psi(\alpha)$ *satisfies* $\lim_{\alpha \to 0} \frac{\psi(\alpha)}{\alpha} = 0$

We know that x^* is a minimum of f, hence this implies that $g(0) = f\left(x^*\right)$. This means that 0 is a minimum of g.

We make the claim that $g'(0) = 0$ because we said that 0 is a minimum of g. To prove this statement we can proceed as follows.

Suppose that $g'(0) \neq 0$. This means that

$$\exists\ \varepsilon > 0,\ \textit{such that}\ \forall\ |\alpha| < \varepsilon\ \textit{and}\ \alpha \neq 0, \left|\frac{\psi(\alpha)}{\alpha}\right| < \left|g'(0)\right| \rightarrow |\psi(\alpha)| < \left|g'(0) \cdot \alpha\right|$$

Hence, from equation 7, we know that

$$g(\alpha) - g(0) < g'(0) \cdot \alpha + |g'(0) \cdot \alpha|$$

If α has an opposite sign to $g'(0)$, we obtain the following:

$$g(\alpha) - g(0) < 0 \rightarrow g(\alpha) < g(0)$$

This is a contradiction since we said that x^* is a minimum for f. Hence, $g'(0) = 0$. At this point, we need to express the result that we obtained above in terms of the function f. If we apply the chain rule from Vector Calculus, we obtain:

$$g'(\alpha) = \nabla f\left(x^* + \alpha \cdot d\right) \cdot d$$

When $\alpha = 0$:

$$g'(0) = \nabla f\left(x^*\right) \cdot d = 0$$

where $\nabla f = \left(f_{x_1}, f_{x_2}, \ldots, f_{x_n}\right)^T$ *represents the gradient of* f

Since d is an arbitrary vector, this implies that $\nabla f\left(x^*\right) = 0$. Now that we have seen this technique, we will apply the same logic to arrive at the First Variation of a Functional.

DEFINITION (FIRST VARIATION)

Let $J : V \to \mathbb{R}$ *be a functional and* $y \in V$.

A linear functional $\delta J|_y : V \to \mathbb{R}$ *is the first variation of* J *at* y *if* $\forall \eta \in V$ *and* $\alpha \in \mathbb{R}$, *we have*

$J(y + \alpha \cdot \eta) = J(y) + \delta J|_y(\eta) \cdot \alpha + o(\alpha)$ **(Equation 7.1)**

The first variation

$$\delta J|_y(\eta)$$

can be seen as the derivative of J in the direction of η at y. Indeed, if we define

$\phi(\alpha) = J(y + \alpha\eta)$, *we have by equation* 7.1, *that* $\delta J|_y(\eta) = \phi'(0) = \lim\limits_{\alpha \to 0} \dfrac{\phi(\alpha) - \phi(0)}{\alpha}$,

therefore, $\delta J|_y(\eta)$ *exists if and only if* ϕ *is differentiable at* 0.

Now, let $y^\star \in V$ be a local minimum for J over $A \subset V$. Consider also the function

$\phi(\alpha) = J\left(y^* + \alpha \cdot \eta\right)$, *for an arbitrary* $\eta \in V$, *such that* $y^* + \alpha \cdot \eta \in A$, $\forall \alpha$ *small enough*

The Taylor expansion for $\phi(\alpha)$ centered at $\alpha = 0$ is given by:

$$\phi(\alpha) = \phi(0) + \phi'(0) \cdot \alpha + \xi(\alpha) \quad \textbf{(Equation 7.2)}$$

$$\text{where } \lim_{\alpha \to 0} \frac{\xi(\alpha)}{\alpha} = 0$$

Since we said that $y^\star \in V$ is a minimum for J over $A \subset V$, and since

$$\phi(0) = J\left(y^*\right)$$

we have that 0 is a minimum of Φ. We make the claim, like in the previous computation that ϕ'(0) = 0 . Analogously like before, we can proceed as follows:

Suppose that $\phi'(0) \neq 0$. This means that

$$\exists\ \varepsilon > 0,\ such\ that\ \forall\ |\alpha| < \varepsilon\ and\ \alpha \neq 0,\ \left|\frac{\xi(\alpha)}{\alpha}\right| < \left|\phi'(0)\right| \rightarrow |\xi(\alpha)| < \left|\phi'(0)\cdot\alpha\right|$$

Hence, from equation C, we know that

$$\phi(\alpha) - \phi(0) < \phi'(0)\cdot\alpha + \left|\phi'(0)\cdot\alpha\right|$$

If α has an opposite sign to $\phi'(0)$, we get the following:

$$\phi(\alpha) - \phi(0) < 0 \rightarrow \phi(\alpha) < \phi(0)$$

This is a contradiction since we said that 0 is a minimum for Φ. Hence, $\phi'(0) = 0$

To better understand why the above result is true, notice that the Taylor Expansion for $J(y^\star + \alpha\eta)$ can be written as:

$$J\left(y^* + \alpha\cdot\eta\right) = J\left(y^*\right) + \phi'(0)\cdot\alpha + o(\alpha)$$

We have thus proved the result.

***THEOREM* (FIRST ORDER NECESSARY CONDITION)**

Let y^* be a local min for the functional J over $A \subset V$. Then, $\delta J\,|_{y^*}(\eta) = 0$ $\forall\,\eta \in V$ such that $y^* + \alpha\cdot\eta \in A$ for α small enough.

We have reached the First Order Necessary Condition. Now that we have seen an interesting method that uses ideas from Optimization to derive the First Variation and the First Order Necessary Condition, we will explore the second Variation as well as the Second Order Necessary Condition.

***DEFINITION* (SECOND VARIATION)**

A quadratic form $\delta^2 J|_y$ *defined from the domain* V *to* $\mathbb{R}$, *is the Second Variation of* J *at* y

if $\forall\, \eta \in V$ *and* α *close to* 0 , *we have*

$$J(y+\alpha\cdot\eta) = J(y) + \delta J|_y(\eta)\cdot\alpha + \delta^2 J|_y(\eta)\cdot\alpha^2 + o(\alpha^2)$$

***THEOREM* (SECOND ORDER NECESSARY CONDITION)**

If y^* *is a local min of* J *over* $A \subset V$, *then* $\forall$ *admissible perturbations* η, *we have*
$\delta^2 J|_{y^*}(\eta) \geq 0$

The statement above is analogous to the Second Order necessary Condition concerning optimization of multivariable real-valued functions, discussed in the previous chapter. More specifically

$$\delta^2 J|_{y^*}(\eta) \geq 0 \ \ and \ \nabla^2 f\left(x^*\right) \geq 0$$

are very similar to each other. At this point, one may think that a second order Sufficient Condition exists also for functionals, as we have seen in the previous chapter concerning optimization, that takes the form

$$\delta^2 J|_{y^*}(\eta) > 0$$

Hence, this implies that $y^\star$ is a minimum point for J. However, it turns out that is **not the case**. The proof of why the statement is false will not be shown in this thesis. The general idea is that the linear increments $J(y^\star + \alpha\eta)$ are too few to guarantee that $y^\star$ is a local minimum.

We are now ready to explore a more explicit formula for the Second Variation of a functional as well as Legendre's condition for a weak minimum. Before moving on, we will consider two examples of computing the First and Second Variation of a functional. These two examples are based on exercise 1.5 and 1.6 taken from the book " Calculus of Variations and Optimal Control Theory: A Concise Introduction" written by Daniel Liberzon.

***EXAMPLE* (First Variation)**

Consider the space $V = \mathbb{C}^0\left(\left[0,1\right], \mathbb{R}\right)$, *let* $\phi: \mathbb{R} \to \mathbb{R}$ *be a* $\mathbb{C}^1$ *function, and define the functional* J *on* V *by* $J(y) = \int_0^1 \phi(y(x))\,dx$.

Show that its first variation exists and is given by the formula $\delta J|_y(\eta) = \int_0^1 \phi'(y(x)) \cdot \eta(x)\,dx$

To solve this exercise, recall that

$$J(y + \alpha \cdot \eta) = J(y) + \delta J|_y(\eta) \cdot \alpha + o(\alpha)$$

In this case $J(y + \alpha\eta)$ is given by $J(y + \alpha\eta) = \int_0^1 \Phi(y(x) + \alpha\eta(x))\,dx$. We will now compute a Taylor expansion for $g(\alpha) = J(y + \alpha\eta)$.

$$J(y + \alpha\eta) = \int_0^1 \phi(y(x))\,dx + \alpha \cdot \int_0^1 \phi'(y(x)) \cdot \frac{d}{d\alpha}(y + \alpha\eta)\,dx + o(\alpha)$$

where $\lim_{\alpha \to \infty} \frac{o(\alpha)}{\alpha} = 0$

$$J(y + \alpha\eta) = \int_0^1 \phi(y(x))\,dx + \alpha \cdot \int_0^1 \phi'(y(x)) \cdot \eta(x)\,dx + o(\alpha)$$

$$J(y + \alpha \cdot \eta) = J(y) + \delta J|_y(\eta) \cdot \alpha + o(\alpha)$$

If we match the two terms highlighted in red above, we obtain the following:

$$\delta J|_y(\eta) = \int_0^1 \phi'(y(x)) \cdot \eta(x)\,dx$$

This is the First Variation of the functional $J(y)$. We also know that the First Variation exists because $\phi: \mathbb{R} \to \mathbb{R}$ *is a* $\mathbb{C}^1$ *function*. The Taylor expansion that we computed above is valid because we know that ϕ is differentiable at 0. Moreover, since

$$g(\alpha) = J(y + \alpha\eta) = \int_0^1 \phi(y(x) + \alpha\eta(x))\,dx$$

we can compute a Taylor expansion for $g(\alpha)$ which is a single variable function, namely $g(\alpha) = g(0) + \alpha\, g'(0)$. The term

$$g(0) = \int_0^1 \phi(y(x))\, dx = J(y)$$

Moreover, the term

$$g'(0) = \frac{d}{d\alpha}\Big|_{\alpha=0} \int_0^1 \phi(y(x) + \alpha\eta(x))\, dx = \int_0^1 \phi'(y(x)) \cdot \eta(x)\, dx = \delta J|_y(\eta)$$

By defining $g(\alpha) = J(y + \alpha\eta)$, we were able to compute the Taylor expansion since $g(\alpha)$ is a function taking as input $\alpha \in \mathbb{R}$.

***EXAMPLE* (Second Variation)**

Consider the space $V = \mathbb{C}^0([0,1], \mathbb{R})$, *and define the functional* J *on* V *by* $J(y) = \int_0^1 \phi(y(x))\, dx$.

Assume now that let $\phi: \mathbb{R} \to \mathbb{R}$ *be a* $\mathbb{C}^2$ *function. Derive the second variation of* J.

The logic to solve this exercise is exactly the same as the one used in the first example. The only difference is that the Taylor expansion will include more terms. We have

$$J(y + \alpha\eta) = \int_0^1 \phi(y + \alpha\eta)\, dx$$

We will now compute the Taylor expansion for $g(\alpha) = J(y + \alpha\eta)$.

$$J(y+\alpha\eta) = \int_0^1 \phi(y(x))\, dx + \alpha \cdot \int_0^1 \phi'(y(x)) \cdot \frac{d}{d\alpha}(y + \alpha\eta)\, dx + \frac{\alpha^2}{2}\int_0^1 \phi''(y(x)) \cdot \left(\frac{d}{d\alpha}(y+\alpha\eta)\right)^2 dx$$

$$J(y+\alpha\eta) = \int_0^1 \phi(y(x))\, dx + \alpha \cdot \int_0^1 \phi'(y(x)) \cdot \eta\, dx + \frac{\alpha^2}{2}\int_0^1 \phi''(y(x)) \cdot \eta^2\, dx$$

Recall that $J(y+\alpha\cdot\eta) = J(y) + \delta J|_y(\eta)\cdot\alpha + \delta^2 J|_y(\eta)\cdot\alpha^2 + o(\alpha^2)$

If we match the two terms highlighted in red above, we obtain:

$$\delta^2 J|_y(\eta) = \frac{1}{2}\int_0^1 \phi''(y(x))\cdot\eta^2(x)\, dx$$

This is the Second Variation for the functional J. The reason why the Taylor expansion shown above is valid is exactly the same as the justification given in example 1.

Now, let's go back to our discussion concerning how to obtain a more explicit expression for the second variation of a functional.

Recall the following expression

if $\forall\,\eta \in V$ *and* α, *we have* $J(y+\alpha\cdot\eta) = J(y) + \delta J|_y(\eta)\cdot\alpha + \delta^2 J|_y(\eta)\cdot\alpha^2 + o(\alpha^2)$

(equation 7.3)

The term given by the following

$$\delta^2 J|_y(\eta)$$

denotes the Second Variation of the functional J. Suppose the functional J takes the following form

$$J(y) = \int_a^b L(x,\, y(x),\, y'(x))\, dx$$

where L is the Lagrangian function, and it is given by $L(x,\, y(x),\, y'(x))$. More specifically,

$$L: \mathbb{R}\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}$$

Moreover, among all $\mathbb{C}^1$ curves $y\colon [a, b] \to \mathbb{R}$ that satisfy $y(a) = y_0$ and $y(b) = y_1$, we want to find local minima of the functional given by

$$J(y) = \int_a^b L(x,\, y(x),\, y'(x))\, dx$$

Hence, we must compute both the first and second variation of a functional. We begin by assuming that $L \in \mathbb{C}^3$. The left hand side of the above equation can be expressed as:

$$J(y+\alpha\cdot\eta)=\int_a^b L(x,\ y(x)+\alpha\cdot\eta(x)\,,\ y'(x)+\alpha\cdot\eta'(x))\ dx \quad where\ L\ is\ the\ Lagrangian$$

We are considering functions $\eta(x)$ such that $\eta(a) = \eta(b) = 0$. We will now construct a Taylor expansion of the above expression with respect to the parameter α. To keep the notation clear and less convoluted we will write L instead of $L(x,\ y(x),\ y'(x))$. Additionally, we will denote by L_y the derivative of L with respect to its second argument evaluated at $y(x)$. The same reasoning applies for the term $L_{y'}$.

$$J(y+\alpha\cdot\eta)=\int_a^b L\,dx + \alpha\cdot\int_a^b\left[L_y\cdot\frac{d}{d\alpha}(y(x)+\alpha\cdot\eta(x))+L_{y'}\cdot\frac{d}{d\alpha}(y'(x)+\alpha\cdot\eta'(x))\right]dx +$$

$$\frac{\alpha^2}{2}\cdot\int_a^b\left[L_{yy}\cdot\left(\frac{d}{d\alpha}(y(x)+\alpha\cdot\eta(x))\right)^2\right]dx + \frac{\alpha^2}{2}\cdot\int_a^b\left[L_{yy'}\cdot\frac{d}{d\alpha}(y'(x)+\alpha\cdot\eta'(x))\cdot\frac{d}{d\alpha}(y(x)+\alpha\cdot\eta(x))\right]dx$$

$$+\frac{\alpha^2}{2}\cdot\int_a^b\left[L_{y'y}\cdot\frac{d}{d\alpha}(y'(x)+\alpha\cdot\eta'(x))\cdot\frac{d}{d\alpha}(y(x)+\alpha\cdot\eta(x)) + L_{y'y'}\cdot\frac{d}{d\alpha}(y'(x)+\alpha\cdot\eta'(x))^2\right]dx$$

$$+\ o(\alpha^2)$$

By computing the derivatives of the above expression and combining like terms, we obtain the following:

$$J(y+\alpha\cdot\eta)=\int_a^b L\,dx + \alpha\cdot\int_a^b\left[L_y\cdot\eta(x) + L_{y'}\cdot\eta'(x)\right]dx + \frac{\alpha^2}{2}\cdot\int_a^b\left[L_{yy}\cdot\eta^2(x)+2\cdot L_{yy'}\cdot\eta(x)\cdot\eta'(x) + L_{y'y'}\cdot(\eta'(x))^2\right]dx$$

$$+\ o(\alpha^2)$$

Matching term by term with **equation 7.3**, we obtain the following:

$$\delta^2 J|_y(\eta) = \frac{1}{2}\cdot\int_a^b\left[L_{yy}\cdot\eta^2(x)+2\cdot L_{yy'}\cdot\eta(x)\cdot\eta'(x) + L_{y'y'}\cdot(\eta'(x))^2\right]dx$$

Let’s now use Integration by parts on the term given by

$$\int_a^b \left[2 \cdot L_{yy'} \cdot \eta(x) \cdot \eta'(x)\right] dx$$

Notice that the term $\int_a^b \left[2 \cdot L_{yy'} \cdot \eta(x) \cdot \eta'(x)\right] dx$ *can be written as* $\int_a^b \left[L_{yy'} \cdot \frac{d}{dx}(\eta(x))^2\right] dx$

The reason is because the perturbation $\eta(x)$ is a function of x. Let's now apply Integration by Parts:

$$\int_a^b \left[L_{yy'} \cdot \frac{d}{dx}(\eta(x))^2\right] dx \text{ , setting } U = L_{yy'} \text{ , } U' = \frac{d}{dx}\left(L_{yy'}\right), \; T' = \frac{d}{dx}(\eta(x))^2 \text{ , } T = \eta^2(x)$$

$$\int_a^b \left[L_{yy'} \cdot \frac{d}{dx}(\eta(x))^2\right] dx = \left[L_{yy'} \cdot \eta^2(x)\right]_a^b - \int_a^b \frac{d}{dx}\left(L_{yy'}\right) \cdot \eta^2(x) \, dx$$

$$\int_a^b \left[L_{yy'} \cdot \frac{d}{dx}(\eta(x))^2\right] dx = -\int_a^b \frac{d}{dx}\left(L_{yy'}\right) \cdot \eta^2(x) \, dx$$

The reason why the expression highlighted above in red simplified in that way is because

$$\eta(a) = \eta(b) = 0 \text{ and hence } \left[L_{yy'} \cdot \eta^2(x)\right]_a^b = 0$$

Hence, the Second Variation can be written explicitly as:

$$\delta^2 J|_y(\eta) = \frac{1}{2} \cdot \int_a^b \left[L_{yy} \cdot \eta^2(x) - \frac{d}{dx}\left(L_{yy'}\right) \cdot \eta^2(x) + L_{y'y'} \cdot (\eta'(x))^2\right] dx$$

$$\text{If we denote } \theta(x) = \frac{1}{2} \cdot L_{y'y'} \quad \text{and } \kappa(x) = \frac{1}{2} \cdot \left(L_{yy} - \frac{d}{dx}\left(L_{yy'}\right)\right)$$

We obtain the following:

$$\delta^2 J|_y(\eta) = \int_a^b \left[\theta(x) \cdot (\eta'(x))^2 + \kappa(x) \cdot (\eta(x))^2\right] dx$$

The functions $\theta(x)$ and $\kappa(x)$ are continuous when $y \in \mathbb{C}^2$.

If y *is a minimum for the functional* J, *then for every* $\mathbb{C}^1$ *perturbations* η *satisfying* $\eta(a) = \eta(b) = 0$

we have $\displaystyle\int_a^b \left[\theta(x) \cdot (\eta'(x))^2 + \kappa(x) \cdot (\eta(x))^2\right] dx \geq 0$ **(equation 7.4)**

Moreover, to arrive at Legendre's condition for a weak minimum, we need to understand the behavior of the functions $\theta(x)$ and $\kappa(x)$. We will not provide the derivation of Legendre's condition for a weak minimum, however, we will just include the statement below.

THEOREM **(Legendre's condition for a weak minimum)**

$\forall\, x \in [a,b]$, *we must have that* $L_{y'y'}(x, y(x), y'(x)) \geq 0$ *for* ***equation*** 7.4 *to hold.*

Building upon Legendre's condition for a weak minimum, we arrive at the Second Order Sufficient Condition. We will not show the proof in this thesis:

THEOREM **(Second Order Sufficient Condition for Optimality)**

An extremal $y(x)$ *is a strict minimum if* $L_{y'y'}(x, y(x), y'(x)) > 0$

$\forall\, x \in [a,b]$ *and the interval* $[a,b]$ *contains no conjugate points to* a.

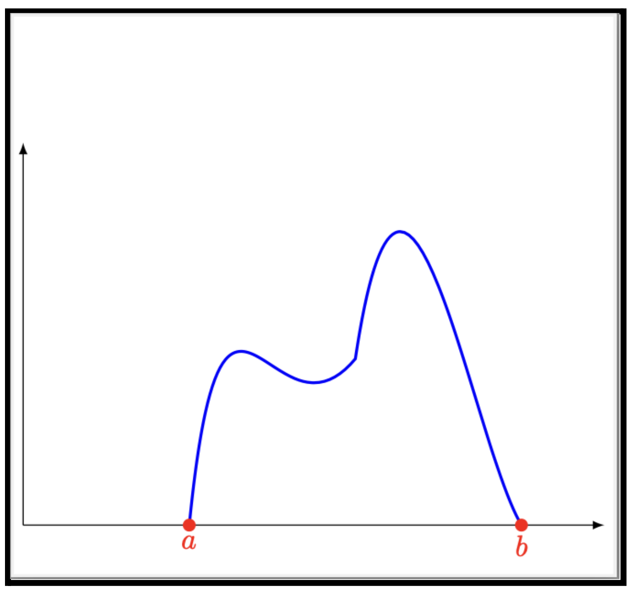
As an important remark, we do not discuss completely in this section of the thesis the Second Order Sufficient Condition for Optimality, but in the specific case in which y is a function of an interval $[a, b]$, we can still have a formulation as given above.

In the next section of the thesis, we will derive the Euler-Lagrange differential equation, which is just an explicit formulation of the First Variation of a function J using two different methods. The rationale of using two different methods to arrive at the same result is to showcase two different ways of reasoning.

Chapter 8: The Euler-Lagrange equation derivation and intuition

There are many different Calculus of Variations problems, for example the Brachistochrone problem, which will be explored in the succeeding chapter of this thesis, the Catenary problem explored by Johann Bernouilli, and Dido's isoperimetric problems. The Euler-Lagrange equation proves to be a valuable tool in solving some of the aforementioned problems. The Euler-Lagrange differential equation is an explicit generalization of the First Order Necessary condition for optimality, discussed in the previous chapter. In this chapter as well as in the succeeding chapters, we will explain the theory behind minimization problems, but of course, all the theory can be reformulated concerning maximization problems.

For example, Dido's isoperimetric problem tries to solve the problem depicted in **Figure 11**:



(Figure 11)

In **figure 11**, we have drawn in black a curve of fixed length. Moreover, the admissible curves are graphs of continuous functions

$$y:[a,b] \rightarrow \mathbb{R}$$

that satisfy $y(a) = y(b) = 0$.

The problem is to find a positive function y such that the area between the x-axis and the graph of y is maximized. More specifically,

$$J(y) = \int_a^b y(x)\,dx$$

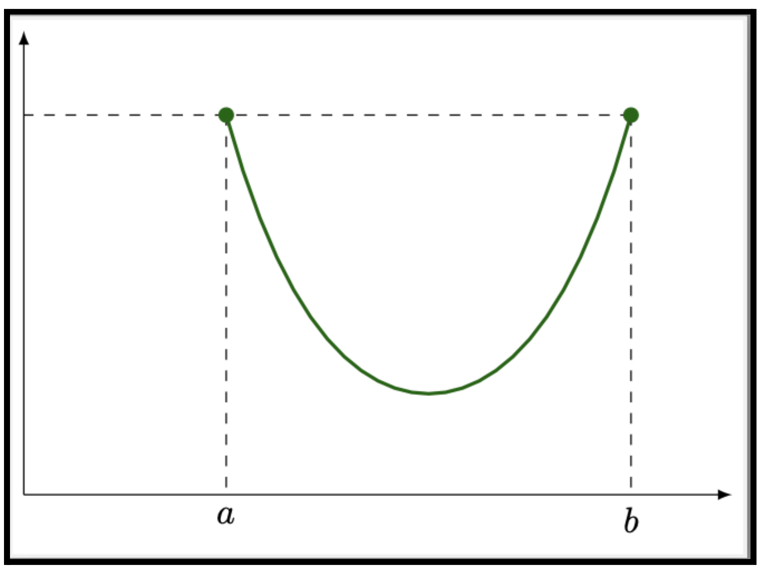
is the functional to be maximized, and the constraint is given by

$$\int_a^b \sqrt{1 + (y'(x))^2}\,dx = C_0 \quad where\ C_0 \in \mathbb{R}$$

See page 85 for a more thorough explanation concerning why this is the correct way to measure the length of the curve. In this case, the constraint embedded into this problem requires that all paths/curves have a constant length.

Another famous Calculus of Variations problem is the Catenary problem.

Figure 12 represents the problem that we are trying to solve



(Figure 12)

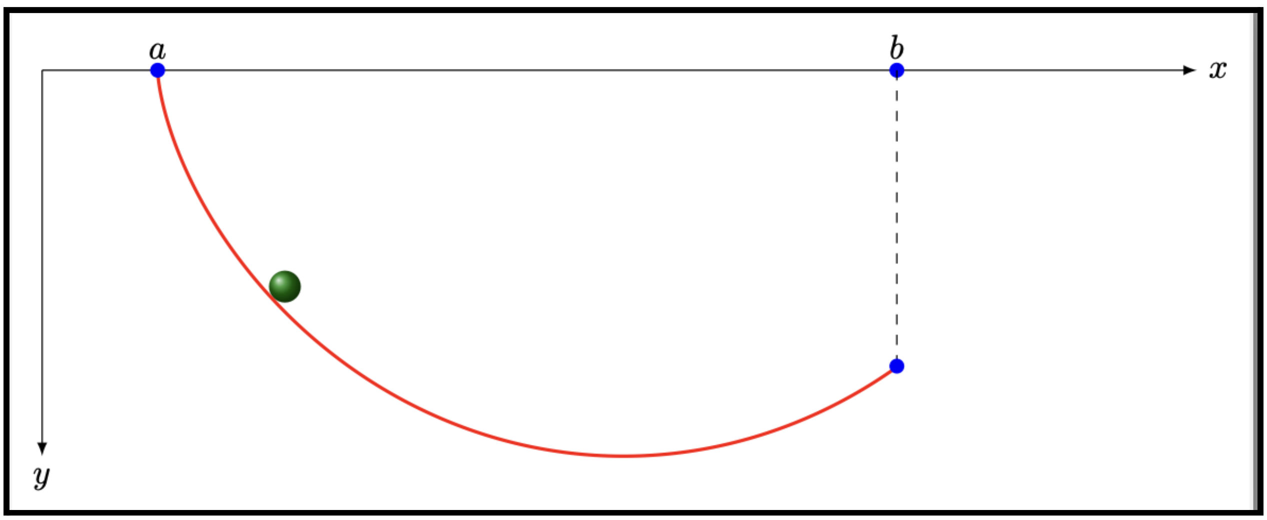
We have a cable of fixed length, fixed at two points a and b. We want to understand what shape the cable makes under the influence of gravity, as shown in figure 12.

The cable takes the shape that minimizes its potential energy. If we model the cable as a function $y(x)$, then we want to

$$\textit{Minimize } J(y) = \int_a^b m \cdot g \cdot y \cdot \sqrt{1 + (y'(x))^2}\, dx \quad \textit{subject to the constraint } \int_a^b \sqrt{1 + (y'(x))^2}\, dx = C_0$$

where $C_0 \in \mathbb{R}$ *and m denotes the mass of the cable and g denotes the acceleration due to gravity*

Lastly, the Brachistochrone is another Calculus of Variations problem, which will be discussed in detail in the next section of this thesis. In the Brachistochrone problem, we are trying to find a function $y(x)$ that defines a curve fixed at two points $A(a,\ y_0)$ and $B(b,\ y_1)$, such that the time for a ball to move from point A to point B along the curve is minimized. **Figure 13** provides a visual representation.



(Figure 13)

The functional that we want to minimize in this case is given by

$$J(y) = \int_a^b \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{2 \cdot g \cdot y}}\, dx$$

A more thorough derivation of this equation will be given in the succeeding chapter of this thesis.

Now that we have seen some basic Calculus of Variation problems, we will move to the derivation of the Euler-Lagrange differential equation.

Consider the function $L(x,\ y(x),\ y'(x))$ where $L: \mathbb{R}\times\mathbb{R}\times\mathbb{R} \ \rightarrow \mathbb{R}$. Among all the possible curves $y:[a,b] \rightarrow \mathbb{R}$ that satisfy the boundary conditions given by $y(a) = y_0$ and $y(b) = y_1$, we want to find the specific function $y(x)$ that minimizes the functional given by

$$J(y) = \int_a^b L(x,\ y(x),\ y'(x))\ dx$$

y is a curve that connects the points $(a,\ y_0)$ and $(b,\ y_1)$. L is the Lagrangian and $y(x)$ and $y'(x)$ denote the position and velocity respectively of a particle moving along the curve.

DERIVATION USING METHOD 1 (Euler – Lagrange equation)

We begin by considering $y+\alpha\cdot\eta$ where η is a perturbation such that $\eta:[a,b] \rightarrow \mathbb{R}\ is\ a\ \mathbb{C}^1$ function, and the parameter α varies in an interval around 0 in $\mathbb{R}$. Moreover, we denote by

$$L_x,\ L_y,\ L_{xx}\ ,\ L_{yx},\ etc$$

the partial derivatives of the Lagrangian function $L(x, y(x),\ y'(x))$. The curve given by $y\ +\ \alpha\eta$ must satisfy $y(a)\ = y_0$ and $y(b)\ =\ y_1$ which means that $\eta(a)\ =\ \eta(b)\ =\ 0$. Note that $y(a)\ = y_0$ and $y(b)\ =\ y_1$ are the boundary conditions of our problem.

Recall that the First Variation satisfies the following expression

$$J(y+\alpha\cdot\eta) = J(y) + \delta J\,|_y\,(\eta)\cdot\alpha\ +\ o(\alpha) \qquad \textbf{(equation 8)}$$

Furthermore, since we know an explicit formula for $J(y)$, we can write equation 8 as

$$J(y+\alpha\cdot\eta) = \int_a^b L(x,\ y(x)+\alpha\cdot\eta(x),\ y'(x)+\alpha\cdot\eta'(x))\ dx$$

Now, by using the Multivariable Chain Rule, we write the Taylor expansion for J around $\alpha\ =\ 0$, obtaining

$$J(y+\alpha\cdot\eta) = \int_a^b \left[L(x,y,y') + L_y(x,y,y')\cdot a\cdot\eta + L_{y'}(x,y,y')\cdot a\cdot\eta' + o(\alpha)\right]dx$$

If we match the above expression with the First Variation expression, depicted in **equation 8**, we obtain the following:

$$\delta J\,|_y(\eta) = \int_a^b \left[L_y(x,y,y')\cdot\eta + L_{y'}(x,y,y')\cdot\eta'\right]dx$$ **(equation 8.1)**

The above expression is an explicit form of the First Variation of a functional J. However, this is not particularly convenient since the second term depends on η'; we say in this case that the first variation is in a weak form. We can apply integration by parts on the second term, to turn the weak form of the First Variation into a strong form.

$$\int_a^b L_{y'}(x,y,y')\cdot\eta'\,dx \quad setting\;\; U = L_{y'}(x,y,y')\;,\; U' = \frac{d}{dx}\left(L_{y'}(x,y,y')\right)\;,\; V' = \eta'\;,\; V = \eta$$

After applying integration by parts, we obtain:

$$\int_a^b L_{y'}(x,y,y')\cdot\eta'\,dx = \left[L_{y'}(x,y,y')\cdot\eta\right]_a^b - \int_a^b \frac{d}{dx}\left(L_{y'}(x,y,y')\right)\cdot\eta\,dx$$

$$\int_a^b L_{y'}(x,y,y')\cdot\eta'\,dx = -\int_a^b \frac{d}{dx}\left(L_{y'}(x,y,y')\right)\cdot\eta\,dx$$

The term

$$\left[L_{y'}(x,y,y')\cdot\eta\right]_a^b$$

evaluates to zero because of the boundary conditions $\eta(a) = \eta(b) = 0$. Hence, after substituting the above expression highlighted in red into **equation 8.1**, we obtain:

$$\delta J\,|_y(\eta) = \int_a^b \left[L_y(x,y,y') - \frac{d}{dx}\left(L_{y'}(x,y,y')\right)\right]\cdot\eta\,dx = 0\;,\;\forall\,\mathbb{C}^1\;curves\;that\;satisfy$$

$$\eta(a) = \eta(b) = 0$$

If y is optimal, that is it solves our minimization problem, the integral evaluates to 0 since we proved in a previous section of the thesis that if y is an optimal curve, then

$$\delta J\,|_y\,(\eta) = 0$$

LEMMA 1

If a continuous function $\psi:[a,b] \to \mathbb{R}$ *is such that* $\int_a^b \psi(x) \cdot \eta \, dx = 0\,,\ \forall\, \mathbb{C}^1$ *functions*

$\eta:[a,b] \to \mathbb{R}$ *that satisfiy the initial boundary conditions* $\eta(a) = \eta(b) = 0$, *then* $\psi(x) = 0$

The proof of Lemma 1 will not be discussed. Hence, this means that for $y(x)$ to be an extremum for the functional $J(y)$, the following equation must hold:

$$L_y(x,y,y') - \frac{d}{dx}\Big(L_{y'}(x,y,y')\Big) = 0$$

$$L_y(x,y,y') = \frac{d}{dx}\Big(L_{y'}(x,y,y')\Big)\,,\ \forall\, x \in [a,b]$$

The equation highlighted in red above is the Euler-Lagrange differential equation. It is possible to extend the equation to cases where $y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$.
Moreover, notice that the right hand side of the Euler-Lagrange differential equation can be written as

$$\frac{d}{dx}\Big(L_{y'}(x,y,y')\Big) = L_{y'x}(x,y,y') + L_{y'y}(x,y,y') \cdot y' + L_{y'y'}(x,y,y') \cdot y''$$

We just used the Multivariable Chain rule to obtain the above expression. It is interesting to notice that there are some second order partial derivatives involving L and y, for example the term

$$L_{y'y'}(x,y,y') \cdot y''$$

This observation may seem to suggest that

$$L,\ y \in \mathbb{C}^2$$

Recall that the weak form of the First Variation given by **equation 8.1** is the following

$$\delta J\,|_y(\eta) = \int_a^b \left[L_y(x,y,y')\cdot\eta + L_{y'}(x,y,y')\cdot\eta'\right]dx$$

We will now apply Integration by parts to the term given by

$$L_y(x,y,y')\cdot\eta$$

$$\int_a^b L_y(x,y,y')\cdot\eta\,dx \quad setting\ U=\eta\ ,\ U'=\eta'\ ,\ V'=L_y(x,y,y')\ ,\ V=\int_a^x L_y(x,y,y')\,dx$$

$$\int_a^b L_y(x,y,y')\cdot\eta\,dx = \left[\eta\cdot\int_a^x L_y(x,y,y')\,dx\right]_a^b - \int_a^b\left[\eta'\cdot\int_a^x L_y(w,y(w),y'(w))\,dw\right]dx$$

Notice that the term in blue above evaluates to 0 because $\eta(a) = \eta(b) = 0$. We just left with the term highlighted in red above. We can now substitute the term in red into **equation 8.1**. Moreover, we changed a bit the notation in the equation highlighted in red above to avoid confusion with the variable x in the integration limits. We changed the input variables of L from $L(x,\ y(x),\ y'(x))$ to $L(w,\ y(w),\ y'(w))$.

$$\delta J\,|_y(\eta) = \int_a^b L_{y'}(x,\ y(x),\ y'(x))\cdot\eta'\,dx - \int_a^b\left[\eta'\cdot\int_a^x L_y(w,\ y(w),\ y'(w))\,dw\right]dx$$

$$\delta J\,|_y(\eta) = \int_a^b\left[L_{y'}(x,\ y(x),\ y'(x))\cdot\eta' - \eta'\cdot\int_a^x L_y(w,\ y(w),\ y'(w))\,dw\right]dx$$

$$\delta J\,|_y(\eta) = \int_a^b\left[L_{y'}(x,\ y(x),\ y'(x)) - \int_a^x L_y(w,\ y(w),\ y'(w))\,dw\right]\cdot\eta'\,dx$$

Hence, if $y(x)$ is a minimum for the functional $J(y)$, we obtain

$$\int_a^b \left[L_{y'}(x, y(x), y'(x)) - \int_a^x L_y(w, y(w), y'(w))\, dw \right] \cdot \eta'\, dx = 0 \quad \text{since } \delta \cdot J|_y(\eta) = 0$$

if y is a minimum.

LEMMA 2

If a continuous function $\xi : [a, b] \to \mathbb{R}$ *is such that* $\int_a^b \xi(x) \cdot \eta'(x)\, dx = 0$, $\forall\, \mathbb{C}^1$ *functions*

$\eta : [a, b] \to \mathbb{R}$ *with* $\eta(a) = \eta(b) = 0$, *then* $\xi(x)$ *is a constant function.*

The proof of Lemma 2 will not be discussed in the thesis. Hence, from Lemma 2, we obtain that along an optimal curve $y(x)$, we have:

$$L_{y'}(x, y(x), y'(x)) = \int_a^x L_y(w, y(w), y'(w))\, dw + C, \text{ where } C \in \mathbb{R}$$

This means that

$$\frac{d}{dx}\left(L_{y'}(x, y, y') \right) \text{ exists and is equal to } L_y(x, y, y')$$

Hence, for the Euler-Lagrange necessary condition, it is enough to assume that

$$y \in \mathbb{C}^1 \text{ and } L \in \mathbb{C}^1$$

Now that we have gained a solid intuition behind the first method of deriving the Euler-Lagrange differential equation,we will look at the same problem discussed above, from a different perspective, to obtain the same exact solution. The logic behind this second method is exactly the same as in example 1, however, the way the proof is carried out is slightly different.

DERIVATION USING METHOD 2 (Euler – Lagrange equation)

Like in the previous proof, we are trying to optimize the following functional

$$J(y) = \int_a^b L(x, y, y')\, dx$$

Moreover, we assume that $y(x)$ is the optimal path, while $\eta(x)$ is an arbitrary path such that $\eta:[a,b] \to \mathbb{R}$ is $\mathbb{C}^1$.

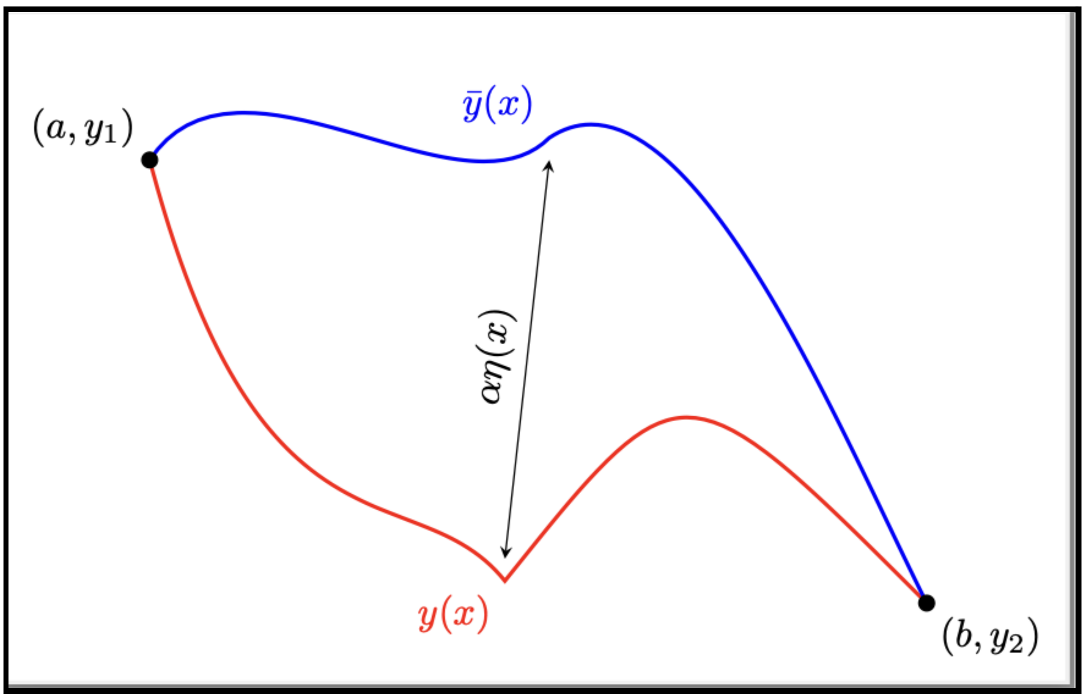
$\bar{y}(x)$ denotes a variation of the function $y(x)$ where α varies in an interval around 0 in $\mathbb{R}$. Now, consider the following expression

$$\bar{y}(x) = y(x) + \alpha \cdot \eta(x)$$

In addition, we require that

$$\bar{y}(a) = y(a) \quad and \; \bar{y}(b) = y(b)$$

This implies that $\eta(a) = \eta(b) = 0$. These are the boundary conditions. **Figure 14** provides a visual representation.



(Figure 14)

Now, we take the derivative with respect to x obtaining

$$\bar{y}'(x) = y'(x) + \alpha \cdot \eta'(x)$$

Recall that the functional in question is given by

$$J(y) = \int_a^b L(x, y, y')\ dx$$

Now, analogously to when we compute the critical points of real-valued functions, we can generalize this idea to functionals. More specifically, given that $y(x)$ is the optimal curve for $J(y)$, this implies that

$$\frac{d}{d\alpha}(J(y)) = 0 \quad as\ \alpha \to 0$$

The reason that we included the condition as $\alpha \to 0$ is because when α approaches 0, we obtain that

$$\bar{y}(x) = y(x)$$

In this case, $y(x)$ is constant since it's the optimal path. On the other hand, $\eta(x)$ is an arbitrary path, but once a particular $\eta(x)$ is chosen, it remains constant. The parameter that changes is thus α. Hence, we obtain

$$\frac{d}{d\alpha}\Big|_{\alpha=0} \int_a^b L\left(x,\ \bar{y}(x),\ \bar{y}'(x)\right) dx = 0$$

By the Leibniz Rule, since the boundaries of integration are constants, we can take the derivative inside the integral:

$$\int_a^b \frac{d}{d\alpha}\left[L\left(x,\ \bar{y}(x),\ \bar{y}'(x)\right)\right]\Big|_{\alpha=0}\ dx = 0$$

Using the Multivariable Chain rule, we obtain:

$$\int_a^b \left[\frac{\partial L}{\partial x}\cdot\frac{dx}{d\alpha} + \frac{\partial L}{\partial \bar{y}} + \frac{d\bar{y}}{d\alpha} + \frac{\partial L}{\partial \bar{y}'}\cdot\frac{d\bar{y}'}{d\alpha}\right]\Big|_{\alpha=0}\ dx = 0$$

$$= \int_a^b \left[\frac{\partial L}{\partial \bar{y}} + \frac{d\bar{y}}{d\alpha} + \frac{\partial L}{\partial \bar{y}'}\cdot\frac{d\bar{y}'}{d\alpha}\right]\Big|_{\alpha=0}\ dx = 0$$

The term $\frac{dx}{d\alpha} = 0$ since x does not depend on α. Moreover, note that

$$\frac{d\bar{y}}{d\alpha} = \eta(x) \quad and \quad \frac{d\bar{y}'}{d\alpha} = \eta'(x)$$

Substituting these values into the above integral expression, we get

$$\int_a^b \left[\frac{\partial L}{\partial \bar{y}} \cdot \eta + \frac{\partial L}{\partial \bar{y}'} \cdot \eta' \right] \Bigg|_{\alpha = 0} dx = 0$$

Recall that as $\alpha \to 0$, the variation of the actual solution approaches the actual solution. In other words

$$As\ \alpha \to 0\ ,\ \bar{y} = y \quad and\ \bar{y}' = y'$$

Hence, when $\alpha = 0$, we obtain

$$\int_a^b \left[\frac{\partial L}{\partial y} \cdot \eta + \frac{\partial L}{\partial y'} \cdot \eta' \right] dx = 0 \qquad \textbf{(equation 8.2)}$$

Now, at this point, we can apply the exact same logic used in method 1 to obtain the Euler-Lagrange differential equation.

In the next chapter we will consider some special cases of the Euler-Lagrange differential equation.

Chapter 9: Special cases of the Euler-Lagrange differential equation

Recall that the Euler-Lagrange differential equation is given by:

$$L_y(x,y,y') = \frac{d}{dx}\Big(L_z(x,y,y')\Big)\ ,\ \forall\, x \in \big[a,b\big]$$

SPECIAL CASE 1

Consider the case when there is no y term in the Lagrangian. The Euler − Lagrange equation reduces to $\frac{d}{dx}\Big(L_{y'}\,(x,\ y')\Big) = 0$

This means that $L_{y'}$ remains constant. Hence, the minima are solutions to

$$L_{y'}\,(x,\ y,\ y') \ = C\ ,\ for\ C \in \mathbb{R}$$

As we will explore recall later on, $L_{y'}\,(x,\ y,\ y')$ is called the momentum.

SPECIAL CASE 2 (**Beltrami Identity**)

Consider the case when the Lagrangian does not depend explicitly on x.

This means that L can be written as $L(y, y')$.

Consider the Euler-Lagrange differential equation again. We will write it this time using a different notation:

$$\frac{\partial L(y,y')}{\partial y} - \frac{d}{dx}\left(\frac{\partial L(y,y')}{\partial y'}\right) = 0$$

We will now multiply both sides by y'.

$$y' \cdot \frac{\partial L(y,y')}{\partial y} - y' \cdot \frac{d}{dx}\left(\frac{\partial L(y,y')}{\partial y'}\right) = 0$$

Now let’s compute the total derivative of $L(y, y')$ with respect to x.

$$\frac{dL(y,y')}{dx} = \frac{\partial L(y,y')}{\partial x} + \frac{\partial L(y,y')}{\partial y} \cdot y' + \frac{\partial L(y,y')}{\partial y'} \cdot y''$$

$$\frac{\partial L(y,y')}{\partial y} \cdot y' = \frac{dL(y,y')}{dx} - \frac{\partial L(y,y')}{\partial x} - \frac{\partial L(y,y')}{\partial y'} \cdot y''$$

Now, by equating the two expressions highlighted in red above, we obtain

$$\frac{dL(y,y')}{dx} - \frac{\partial L(y,y')}{\partial x} - \frac{\partial L(y,y')}{\partial y'} \cdot y'' = y' \cdot \frac{d}{dx}\left(\frac{\partial L(y,y')}{\partial y'}\right)$$

$$\frac{dL(y,y')}{dx} - \frac{\partial L(y,y')}{\partial x} - \left[\frac{\partial L(y,y')}{\partial y'} \cdot y'' + y' \cdot \frac{d}{dx}\left(\frac{\partial L(y,y')}{\partial y'}\right)\right] = 0$$

$$\frac{dL(y,y')}{dx} - \frac{\partial L(xy,y')}{\partial x} - \left[\frac{d}{dx}\left(y' \cdot \frac{\partial L(y,y')}{\partial y'}\right)\right] = 0$$

$$\frac{d}{dx}\left(L(y,y') - y' \cdot \frac{\partial L(y,y')}{\partial y'}\right) = \frac{\partial L(y,y')}{\partial x}$$

Since we said initially that L does not depend explicitly on x , we know that

$$\frac{\partial L(y,y')}{\partial x} = 0$$

Hence, the expression highlighted in pink above becomes

$$\frac{d}{dx}\left(L(y,y') - y' \cdot \frac{\partial L(y,y')}{\partial y'}\right) = 0$$

which implies that

$$L(y,y') - y' \cdot \frac{\partial L(y,y')}{\partial y'} = C \ , \ for \ C \in \mathbb{R}$$

We can multiply both sides by -1 to obtain

$$y' \cdot \frac{\partial L(y,y')}{\partial y'} - L(y,y') = B \ , \ for \ B \in \mathbb{R} \ where \ B = -C$$

An equivalent expression for the above equation using a different notation is

$$L_{y'} \cdot y' - L = B \ , \ for \ B \in \mathbb{R}$$

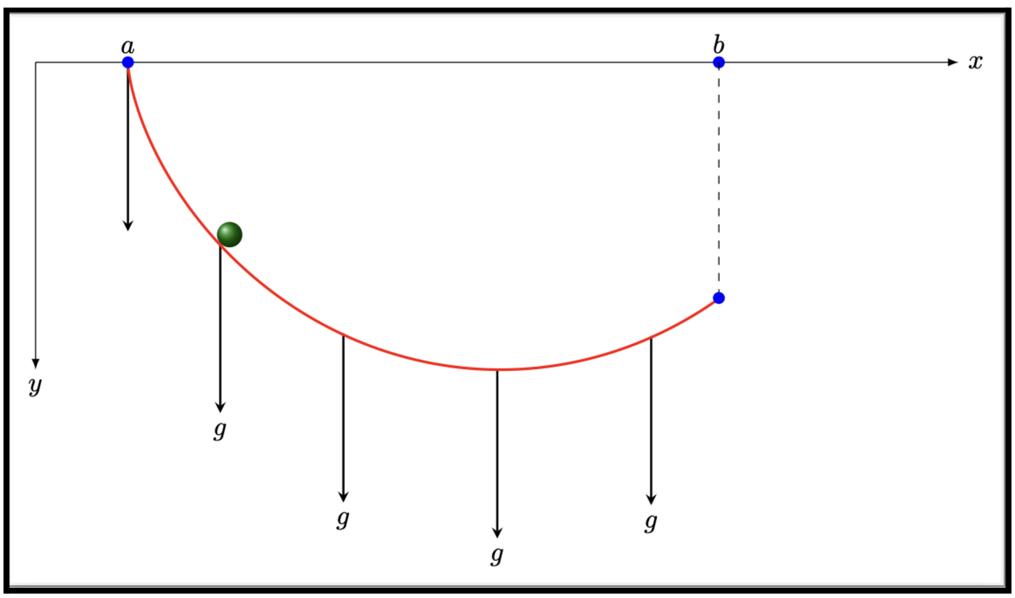
As we will discuss later on, the quantity given by

$$L_{y'} \cdot y' - L = B \ , \ for \ B \in \mathbb{R}$$

is called the Hamiltonian. Its significance will be thoroughly explained in a later chapter.

Chapter 10: The Brachistochrone Problem

We will now solve a very famous Calculus of Variations problem called the Brachistochrone problem using the previously derived Euler-Lagrange differential equation. As a small historical background to this infamous Calculus of Variations problem, in 1670, Calculus was invented by Newton and Leibniz independently. Next, in 1696, Bernoulli proposed the so-called Brachistochrone problem. Later on in 1733 Leonard Euler elaborated on the Brachistochrone problem. Johann Bernouilli's aim of posing this problem to the mathematical community was to "gain the gratitude of the whole scientific community by placing before the finest mathematicians of his time a problem which will test their methods and the strength of their intellect." (Johann Bernouilli)[1]. Galileo incorrectly stated that the path of fastest descent would be an arc of a circle. Newton, Jacob Bernouilli, Leibniz and De L'Hôpital proposed solutions to the problem, and in 1760, Lagrange published an essay concerning new methods of determining the maxima and minima of indefinite integral formulas. [2] Let's now begin to analyze the Brachistochrone problem. Let's consider **figure 15**



(figure 15)

The basic idea is that we are given two fixed points A= $(a\ ,\ y_1)$ and B= $(b\ ,\ y_2)$. Moreover, we want to find a specific curve $y(x)$ where $y:[a,b]\to\mathbb{R}$ such that the time for a ball to move from point a to point b is minimized. Moreover, at each infinitesimal point on the curve, there is a downwards force that acts on the ball, namely the force due to gravity, denoted by g . Let's now formulate the Brachistochrone problem in terms of a Calculus of Variations problem.

[1] J J O'Connor and E F Robertson, The Brachistochrone problem
[2] J J O'Connor and E F Robertson, The Brachistochrone problem

Assume that at time $t = 0$ the ball is at point A. We want to minimize the total time it takes the ball to reach point B. Hence, we can use a definite integral, namely

$$J(y) = \int_0^t dt$$

Recall from Kinematics that $speed = \frac{distance}{time}$. This implies that $time = \frac{distance}{speed}$.

We can rewrite the integral as

$$J(y) = \int_a^b \frac{dS}{v} \quad \textbf{(equation } 10)$$

where dS represents infinitesimally small lengths of the curve and v denotes the velocity of the ball. We use the integral since we want to take the sum of all these infinitesimally small segments of the curve divided by the velocity of the ball at each point on the curve. Using the Pythagorean Theorem, we know that

$$(dx)^2 + (dy)^2 = (dS)^2$$

$$dS = \sqrt{(dx)^2 + (dy)^2}$$

$$dS = \sqrt{(dx)^2 \cdot \left[1 + \frac{(dy)^2}{(dx)^2}\right]}$$

$$dS = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

$$dS = \sqrt{1 + (y')^2}\, dx$$

We now need to figure out an expression for v(the velocity of the ball at each infinitesimally small length of the curve). From the Principle of conservation of energy, we know that when the ball starts at point A it has potential energy, and then the ball gains kinetic energy. By the principle of conservation of energy, this implies that

$$m \cdot g \cdot y = \frac{1}{2} \cdot m \cdot v^2$$

$$v = \sqrt{2gy}$$

Notice that y denotes the height of the ball at each point on the curve. Substituting the two terms highlighted in red above into **equation** 10**,** we obtain:

$$J(y) = \int_a^b \frac{\sqrt{1+(y')^2}}{\sqrt{2gy}} dx$$

The above expression can be rewritten as

$$J(y) = \frac{1}{\sqrt{2g}} \int_a^b \sqrt{\frac{1+(y')^2}{y}}\, dx$$

Hence, we are trying to minimize the functional depicted above. The function L in this case is given by $L(y, y') = \sqrt{\frac{1+(y')^2}{y}}$. Notice that it does not depend on x. We can use the Beltrami identity which corresponds to **special case 2** of the Euler-Lagrange equation; see page 80.

$$L - y' \cdot L_{y'} = K, \text{ for } K \in \mathbb{R}$$

In this case we obtain

$$\sqrt{\frac{1+(y')^2}{y}} - y' \cdot \frac{1}{\sqrt{y}} \cdot \frac{1}{2} \cdot \left(1+(y')^2\right)^{-\frac{1}{2}} \cdot 2 \cdot y' = K$$

$$\sqrt{\frac{1+(y')^2}{y}} - \frac{(y')^2}{\sqrt{y \cdot \left(1+(y')^2\right)}} = K$$

Multiplying both sides by $\sqrt{y}$, we get

$$\sqrt{1+(y')^2} - \frac{(y')^2}{\sqrt{1+(y')^2}} = K\sqrt{y}$$

Multiplying both sides by $\sqrt{1 + (y')^2}$ gives us

$$1+(y')^2 - (y')^2 = K\sqrt{y\cdot\left(1+(y')^2\right)}$$

$$1 = K\sqrt{y\cdot\left(1+(y')^2\right)}$$

Finally, after squaring both sides, the expression reduces to

$$1 = K^2\cdot y\cdot\left(1+(y')^2\right)$$

$$\frac{1}{K^2} = y\cdot\left(1+(y')^2\right)$$

If we denote $\frac{1}{K^2}$ as C_1, we get

$$C_1 = y + y\cdot(y')^2$$

$$\sqrt{\frac{C_1 - y}{y}} = \frac{dy}{dx}$$

Notice that the above differential equation is autonomous and separable.

$$\int dx = \int \sqrt{\frac{y}{C_1 - y}}\ dy$$

$$x + C_2 = \int \sqrt{\frac{y}{C_1 - y}}\ dy$$

Now, we will utilize a trigonometric substitution.

$$y = C_1\cdot \sin^2(\theta)\ ,\ dy = 2\cdot C_1\cdot \sin(\theta)\cdot\cos(\theta)\ d\theta\ ,\ 0 < \frac{\pi}{2} < \theta$$

After applying this substitution, we obtain

$$x + C_2 = \int \sqrt{\frac{C_1 \cdot \sin^2(\theta)}{C_1 - C_1 \cdot \sin^2(\theta)}} \cdot 2C_1 \cdot \sin(\theta) \cdot \cos(\theta)\ d\theta$$

$$= \int \sqrt{\frac{C_1 \cdot \sin^2(\theta)}{C_1 \cdot \cos^2(\theta)}} \cdot 2C_1 \cdot \sin(\theta) \cdot \cos(\theta)\ d\theta$$

$$= \int \tan(\theta) \cdot 2C_1 \cdot \sin(\theta) \cdot \cos(\theta)\ d\theta$$

$$= 2C_1 \int \sin^2(\theta)\ d\theta$$

$$= 2 \cdot C_1 \int \frac{1}{2} \cdot (1 - \cos(2\theta))\ d\theta$$

$$= C_1 \int (1 - \cos(2\theta))\ d\theta$$

$$= C_1 \cdot \left(\theta - \frac{1}{2} \cdot \sin(2\theta)\right)$$

Hence, the solutions are given by

$$x = C_1 \cdot \left(\theta - \frac{1}{2} \cdot sin(2\theta)\right) - C_2$$

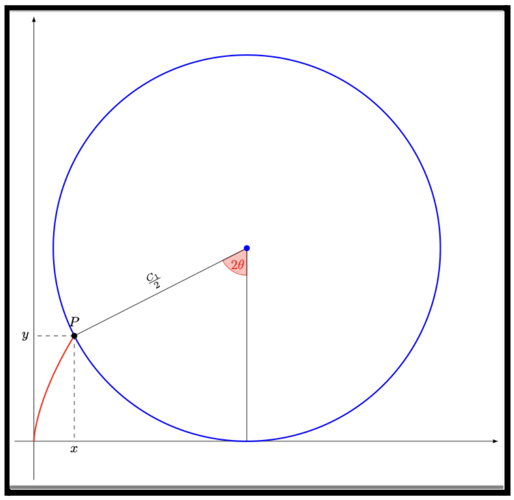
$$y = C_1 \cdot sin^2(\theta) = \frac{C_1}{2} \cdot (1 - cos(2\theta))$$

To find the values of the constants, we can use the initial boundary conditions depending on the situation. The two parametric equations that we obtained above

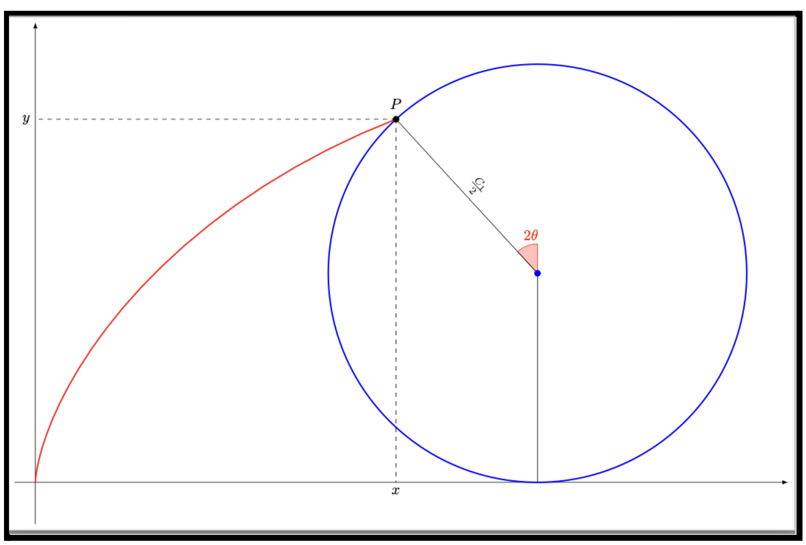
$$x = C_1 \cdot \left(\theta - \frac{1}{2} \cdot \sin(2\theta) \right) - C_2$$

$$y = \frac{C_1}{2} \cdot (1 - \cos(2\theta))$$

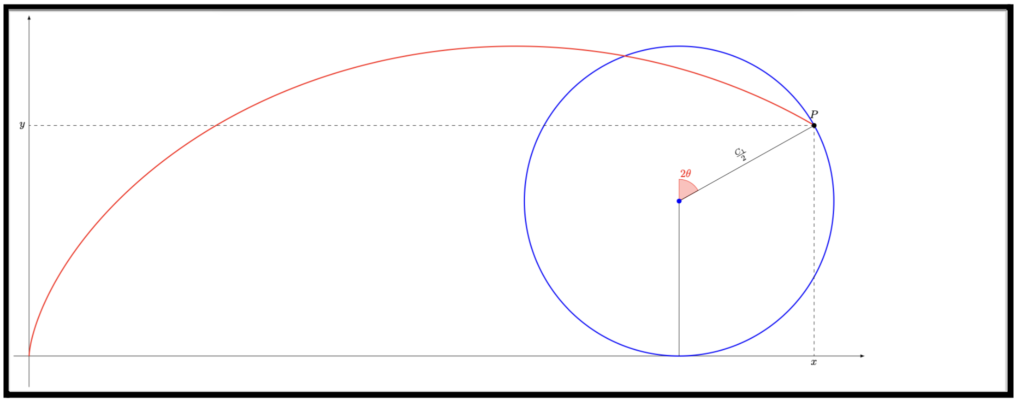
are the equations that define a Cycloid. Let's gain some intuition regarding the shape of a Cycloid. **Figures 16,17,18,19** provide a visual representation.



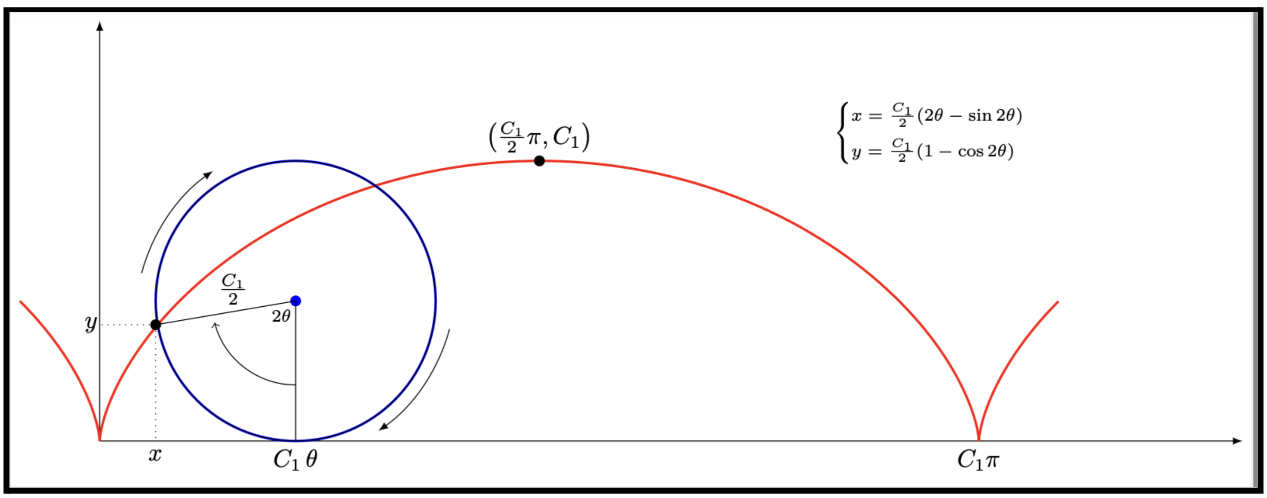
(Figure 16)



(Figure 17)



(Figure 18)



(Figure 19)

From the diagrams above, we can see that a Cycloid is the path traced out by a fixed point on the circumference of a circle as the circle rolls along a line. Hence, relating back to the Brachistochrone problem, using techniques from Calculus of Variations, we figured out that the Cycloid that passes through points $(a,\ y_1)$ and $(b,\ y_2)$ minimizes the time it takes for the ball to move from A to B under the influence of gravity.

Chapter 11: The Hamiltonian

In the previous chapter of the thesis, we applied the Euler-Lagrange equation (Beltrami identity) to solve the famous Brachistochrone problem. In this section of the thesis, we will discuss in more detail the Hamiltonian, the Principle of Least Action as well as the principles of conservation of energy and momentum. All these fundamental ideas in Physics can be formulated using the Hamiltonian and the Euler-Lagrange differential equation.

Suppose we are given a minimization problem with Lagrangian L. The momentum associated to a given $y = y(x)$ is defined as

$$P = L_{y'}(x, y, y')$$

where L is the Lagrangian. The Hamiltonian, for our minimization problem, is defined as

$$H(x, y, y', P) = P \cdot y' - L(x, y, y')$$

In this case, y and P are called canonical variables. Let's now explore Hamilton's canonical equations. Let y be a minimum for our problem that satisfies the Euler-Lagrange differential equations. We know that

$$\frac{dy}{dx} = y'(x) = H_P(x, y, y')$$

Let's now figure out a similar expression for $\frac{dP}{dx}$:

$$\frac{dP}{dx} = \frac{d}{dx}\left(L_{y'}(x, y, y')\right) = -H_y(x, y, y')$$

The equations highlighted in red above are Hamilton's canonical equations, therefore, another necessary condition for having an optimal solution is strictly related to the Hamiltonian. The condition states that H has a stationary point as a function of $y'(x)$ along the optimal curve $y(x)$. To better understand this statement, note that

$$H_{y'}(x, y, y', P) = P - L_{y'}(x, y, y') = 0 \quad \text{since } P = L_{y'}(x, y, y')$$

Moreover, let $x \in [a, b]$ and let y be the optimal curve. We define

$$P(x) = L_{y'}(x, y(x), y'(x))$$

Additionally, the Hamiltonian is given by

$$P(x) = L_{y'}(x, y, y') \cdot y' - L(x, y, y')$$

Consider now $H(z) = L_{y'}(x, y, y')\, z - L(x, y, z)$. One can prove that $H(z)$ has a stationary point when $z = y'(x)$, which represents the velocity of the optimal curve y at x. More specifically, $\frac{dH}{dz}(y'(x)) = 0$.

Another interesting idea that emerges is the Principle of Least Action. Recall Newton's law that states that $F = ma$ where F denotes the resultant force acting on a system, m denotes the mass, and a denotes the acceleration. We can formulate this law as

$$\frac{d}{dt}\left(m\dot{q}\right) = -U_q \quad \textbf{(equation 11)}$$

where m is the mass, $\dot{q} = \frac{dq}{dt}$ denotes the velocity, and finally $U = U(q)$ represents the potential energy.

In addition, $m\dot{q}$ is the momentum, and $-U_q$ is the force acting on a system. It turns out that there is a very nice connection between Newton's law of motion given by equation 11

and the Euler-Lagrange equation. To see this interesting connection, we must rewrite the Euler-Lagrange differential equation in terms of this new notation. More specifically,

$$L(x, y, y') \ \textit{becomes} \ L\left(t, q, \dot{q}\right)$$

Hence, the Euler-Lagrange equation becomes using this new notation:

$$\frac{d}{dt}\left(L_{\dot{q}}\right) = L_q \quad \textbf{(equation 11.1)}$$

The curves are now parameterized with respect to time (t). The central question that we now want to answer is what Lagragian function makes **equation** 11 and **equation** 11.1 the same. If in $\mathbb{R}^3$ we consider

$$L = \frac{1}{2} \cdot m \cdot \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - U(q)$$

$$= \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 - U(q)$$

Equation 11 and **equation** 11.1 are the same. To see why notice that

$$L_{\dot{q}} = m \cdot \dot{q}$$

$$L_q = -U_q$$

The term given by

$$\frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2$$

denotes the kinetic energy, while the term given by $U(q)$ denotes the potential energy. The specific Lagrangian function given by

$$L = \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 - U(q)$$

gives Hamilton's Least Action Principle.

More specifically, the paths of mechanical systems are the extremals of the following integral

$$\int_{t_0}^{t_1} (T - U)\, dt$$

where T denotes the kinetic energy and U denotes the potential energy. Hamilton's Least Action Principle states that a particle will take the path that minimizes the difference between its kinetic and potential energies. If we recall the definition of the Hamiltonian, using the Lagrangian function given by

$$L = \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 - U(q)$$

we arrive at a very intuitive physical interpretation of the Hamiltonian.

$$H = \dot{q} \cdot L_{\dot{q}} - L = \dot{q} \cdot \left[m \cdot \dot{q}\right] - \left[\frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 - U(q)\right]$$

$$= m \cdot \left(\dot{q}\right)^2 - \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 + U(q)$$

$$= \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 + U(q)$$

$$= E_{total}$$

The Hamiltonian thus represents the total energy in a system. We can apply what we discovered about the physical interpretation of the Hamiltonian to the two special cases of the Euler-Lagrange differential equation, to obtain the principles of conservation of energy and momentum. In each one of the cases discussed below, the Lagrangian function L is given by

$$L = \frac{1}{2} \cdot m \cdot \left(\dot{q}\right)^2 - U(q)$$

The conservation of energy principle is based on the **special case 2** of the Euler-Lagrange differential equation, also known as the Beltrami identity.

CONSERVATION OF ENERGY PRINCIPLE

Suppose $L = L\left(q, \dot{q}\right)$

This case corresponds to the $L(y, y')$ *case that we explored in a previous section of this thesis.*

From the **SPECIAL CASE 2 (Beltrami identity)** on page 85, *we know that*

$y' \cdot L_{y'} - L = C\ ,\ C \in \mathbb{R}$

Using this new notation, we can write the Beltrami identity as $\dot{q} \cdot L_{\dot{q}} - L = = H = C\ ,\ C \in \mathbb{R}$

This means that the Hamiltonian is conserved, which is another way of saying that the total energy of a system is conserved since $H = E_{total} = C\ ,\ C \in \mathbb{R}$

Moreover, the conservation of momentum principle is based on **special case 1** of the Euler-Lagrange differential equation.

CONSERVATION OF MOMENTUM PRINCIPLE

Assume that no external force acts on a system. Hence, $L_q = -U_q$ *which is constant. The force* U_q *is constant. The kinetic energy* T *of a system depends on* $\dot{q}$ *and not on* q*. Hence,* $L = T - U$ *does not explicitly depend on* q.

From the **SPECIAL CASE 1**, *given on page* 85, *we know that* $L_{y'} = a\ ,\ a \in \mathbb{R}$.

We can rewrite this statement using our new notation as $L_{\dot{q}} = a\ ,\ a \in \mathbb{R}$.

We also know an explicit expression for $L_{\dot{q}}$ *which equals* $m \cdot \dot{q}$*. This means that momentum is conserved.*

In the next chapter of the thesis we will consider some Calculus of Variations problems dealing with constraints. We will consider just a few situations and we will provide some intuition behind how to solve them.

Chapter 12: Integral, non-integral constraints and variable end point problems

In this last chapter of the thesis, we will consider three different possible plausible scenarios that may emerge when dealing with a Calculus of Variations problem. We will begin with variable end point problems and then we will move on to integral and non-integral constraints.

So far, we considered, as variations, a class of functions denoted by $\eta(x)$, which must equal zero at the respective endpoints. More specifically, $\eta(a) = \eta(b) = 0$. We will now consider a slightly different problem. Consider the functional given by

$$J(y) = \int_a^b L(x, y, y')\ dx \quad where\ y(a) = y_0 \ \ and\ \ y(b)\ is\ free$$

In this case $y(b)$ can take on any value. This means that $\eta(a) = 0$ but $\eta(b)$ is not necessarily equal to zero; it can be equal zero, but it is not restricted to equal zero. Moreover, recall that the First Variation in its explicit form is given by

$$\delta J|_y(\eta) = \int_a^b \left[L_y(x, y, y') - \frac{d}{dx}\Big(L_{y'}(x, y, y') \Big) \right] \cdot \eta\ dx + \Big[L_{y'}(x, y, y') \cdot \eta \Big]_a^b$$

$$\delta J|_y(\eta) = \int_a^b \left[L_y(x, y, y') - \frac{d}{dx}\Big(L_{y'}(x, y, y') \Big) \right] \cdot \eta\ dx + L_{y'}(b, y(b), y'(b)) \cdot \eta(b)$$

Moreover, we know that if y is an extremum for the functional given by $J(y)$, then

$$\delta J|_y(\eta) = 0$$

which also means that

$$\int_a^b \left[L_y(x, y, y') - \frac{d}{dx}\Big(L_{y'}(x, y, y') \Big) \right] \cdot \eta \, dx = 0$$

Finally, this implies that

$$L_{y'}(b, y(b), y'(b)) \cdot \eta(b) = 0$$

since $\eta(b)$ is arbitrary. The equation highlighted in blue above replaces the boundary condition $y(b) = y_1$.

In addition, we may also come across situations in which we have an integral constraint. For example, the Catenary problem defined by

$$\textit{Minimize } J(y) = \int_a^b m \cdot g \cdot y \cdot \sqrt{1 + (y'(x))^2} \, dx \quad \textit{subject to the constraint} \int_a^b \sqrt{1 + (y'(x))^2} \, dx = C_0$$

$$\textit{where } C_0 \in \mathbb{R}$$

is an example of a Calculus of Variations problem with an integral constraint. We can build upon the general problem of optimizing a functional $J(y)$ by introducing a constraint integral:

$$C(y) = \int_a^b M(x, y(x), y'(x)) \, dx = C_0, \; C_0 \in \mathbb{R}$$

However, some complications emerge when the curve y is an extremal of the constraint function C. To rule out degenerate cases, it is imperative to assume that y is not an extremal of C so that there exist nearby curves where C takes on values greater than or smaller than C_0. This is very similar to the non-degeneracy requirement concerning Lagrange multipliers.

Notice that M is a function that belongs to the same class as L. Now, we assume that y is an extremum for $J(y)$. Consider $y + \alpha\eta$. Notice that for $\eta(x)$ to be admissible it has to satisfy the constraint and satisfy that $\eta(a) = \eta(b) = 0$.

We know that $C(y + \alpha\eta) = C_0$, $\forall\, \alpha$ close to 0. This means that

$$\delta C\,|_y(\eta) = 0$$

If we follow the same steps used to derive the Euler-Lagrange differential equation, by following the logic of **METHOD 1**, we get

$$\int_a^b \left[M_y(x,y,y') - \frac{d}{dx}\left(M_{y'}(x,y,y') \right) \right] \cdot \eta(x)\ dx = 0$$

(equation 12**)**

From the First Order Necessary Condition, we know that for every $\eta(x)$ that satisfy **equation** 12 **,** we have

$$\delta J|_y(\eta) = \int_a^b \left[L_y(x,y,y') - \frac{d}{dx}\left(L_{y'}(x,y,y') \right) \right] \cdot \eta(x)\ dx = 0$$

(equation 12. 1**)**

If we pay close attention to **equation** 12 and **equation** 12. 1 **,** we can notice that there is a clear connection with Lagrange Multipliers. More specifically,

$$\exists\ a\ \lambda^* \in \mathbb{R}\ , such\ that \left(L_y - \frac{d}{dx}\left(L_{y'} \right) \right) + \lambda^* \cdot \left(M_y - \frac{d}{dx}\left(M_{y'} \right) \right) = 0\ ,\quad \forall\, x \in [a,b]$$

If we rearrange the terms, we obtain the following expression:

$$\left(L + \lambda^* \cdot M \right)_y = \frac{d}{dx}\left(L + \lambda^* \cdot M \right)_{y'}$$

The term given by

$$L + \lambda^* \cdot M$$

is called the augmented Lagrangian. Hence, y being an extremal for the cost functional $C(y)$ means that

$$\left(J + \lambda^* C \right)(y) = \int_a^b \left[L(x,y,y') + \lambda^* \cdot M(x,y,y') \right] dx$$

In the above section, we considered integral constraints since the cost functional $C(y)$ was written in terms of a definite integral. There are also non-integral, which the name implies, where the cost functional is not written in terms of a definite integral. The difference between integral and non-integral constraints is that integral constraints are global in the sense that they apply to the entire curve, and non-integral constraints are local.

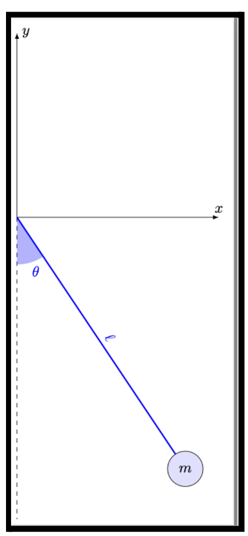
We now consider an equality constraint that has to hold pointwise. That is

$M(x,\ y(x),\ y'(x)) = 0,\ \forall\, x \in [a, b]$. The First Order Necessary condition is very similar to the one that we saw above, concerning integral constraints, the only difference is that $\lambda^*(x)$ is now a function of x. The Euler-Lagrange differential equation must hold for

$$\left(L + \lambda^*(x) \cdot M\right)$$

where $\lambda^* [a, b] \rightarrow \mathbb{R}$. To rule out degenerate cases, we need to make two assumptions. We assume that there are at least 2 degrees of freedom, and that along the curve y, $M_{y'} \neq 0$ or if y' is not present in $M(x, y, y')$ then $M_y \neq 0$. We will not prove it here, but $\forall\ x$, $\exists$ a Lagrange Multiplier where $\lambda^\star = \lambda^\star(x)$. The individual Lagrange Multipliers $\lambda^\star(x)$ combined together generate the function $\lambda^\star$. The problem becomes of minimizing $\int_a^b L\, dx + \int_a^b \lambda(x)\, M\, dx$.

The last type of non-integral constraints that will be explored in this thesis are the Holonomic constraints. Holonomic constraints refer to cases when the constraint does not depend on $y'(x)$, so they take form $M(x,\ y(x)) = 0$. Let's consider an interesting example. Consider **figure 20**.



(Figure 20)

Figure 20 describes a pendulum, having a ball of mass m hanging from a string of fixed length l. Moreover, θ is the angle formed between the string and the vertical y-axis. We will apply Hamilton's least action principle to figure out an equation that describes the motion of the ball subject to the pendulum, such that the difference between the ball's kinetic and potential energy is minimized. Before moving on to the example, we must define what Polar Coordinates are.

***DEFINITION* (Polar Coordinates)**

To move from Cartesian Coordinates (x, y) *to Polar Coordinates* (r, θ) *we apply the following equations*: $x = r \cdot \cos(\theta)$ *and* $y = r \cdot \sin(\theta)$.

In this case, r *is the distance from the origin to the point in Cartesian coordinates*. $x^2 + y^2 = r^2$.

Moreover, $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ *is the angle formed between the* $x-$ *axis and the point given in Cartesian Coordiantes*.

$\theta = arc\tan\left(\frac{y}{x}\right)$

Now let's go back to our initial problem. The constraint function is given by $M(x, y) = x^2 + y^2 - l^2 = 0$ by the Pythagorean Theorem. We can express everything in terms of Polar Coordinates. Note that $r = l$ and θ is a free parameter. Moreover, let's denote by $\hat{\theta} = \frac{d\theta}{dt}$. The kinetic energy as the mass m swings back in forth on the pendulum can be written as $T = \frac{1}{2}ml^2 (\hat{\theta})^2$ and the potential energy can be expressed as $U = mg(L - y)$ which is equivalent to $mg(1 - \frac{y}{l}) = mg(1 - cos(\theta))$. The Lagrangian, by following Hamilton's Least Action Principle, is given by $L = \frac{1}{2}ml^2 - mg(1 - cos(\theta))$. If we apply the Euler-Lagrange differential equation for Polar Coordinates, given by $L_\theta = \frac{d}{dt}(L_{\hat{\theta}})$, we arrive at the famous pendulum equation given by

$$\frac{d^2\theta}{dt^2} = -\frac{g}{l} \cdot \sin(\theta)$$

where g represents the acceleration due to gravity, given by $9.81\, ms^{-2}$, and l represents the length of the string.

Conclusions

In conclusion, in this thesis we started with exploring some general theory behind Differential Equations. We started off with giving the general form of an Ordinary Differential Equation. Next, we explored Autonomous, separable, and finally Second Order linear homogeneous and nonhomogeneous differential equations with constant coefficients. We then gave a brief definition of what it means for a function to be Lipschitz and we provided the general statement of the Picard-Lindelöf theorem. Before proving this theorem, we explored some interesting concepts ranging from Cauchy Sequences, and Cauchy sequences of functions as well as pointwise and uniform convergence.

We then provided some intuition behind the concept of Optimization for real-valued functions having n variables. We also explored the concept of Lagrange Multipliers. Finally, we explored the concept of a Functional,which is the central building block of Calculus of Variations. We then introduced the First and Second Variation of a functional in both a general and explicit form.. We then derived the Euler-Lagrange differential equation and applied it to the infamous Brachistochrone problem. We concluded the thesis by looking at some possible integral and non-integral constraints.

Calculus of Variations is a very important topic since it has a wide range of applications in Physics, Chemistry, etc. Calculus of Variations is analogous to the idea of minimizing or maximizing a function of real variables. The only difference is that in this case, we are trying to find a specific curve (function) that optimizes a functional (which is a function of a function).

As we saw in this thesis, Newton's laws of motion have a clear connection with the Euler-Lagrange differential equation. For further research, it would be very interesting to explore the connection between the Euler-Lagrange differential equation and relativistic systems. The Euler-Lagrange differential equation can be applied in the context of both special and general relativity. Moreover, since the theory of Relativity by Einstein is based upon Differential Geometry, this also implies that Lagrangian Mechanics is connected to Differential Geometry.

Much of the theory that was developed in this thesis can also be applied to Computer science, more specifically, in Machine Learning for example. A Calculus of Variations methodology can be applied to provide a lower bound for the marginal likelihood function [3]. Moreover, Calculus of Variation techniques can also be applied to image restoration, which is based on minimizing a specific functional [4]. For example, image restoration, which is a ML topic, is often based on optimizing

$$I(u) = \int_U \frac{1}{2}(f-u)^2 + \lambda \cdot \left|\nabla u\right| dx$$ [5]

[3] Anders Meng, *An Introduction to Variation Calculus in Machine Learning*
[4] Jeff Calder, *The Calculus of Variations*
[5] ibid

In this case,

$u : U \to \mathbb{R}$ *where* $U = (0,1)^2$. *Moreover*, f *is the original noisy image*, *and* u *denotes the minimizer*, *which represents the denoized image*.

In terms of a Calculus of Variations problem, the Lagrangian is given by:

$$L(x, z, p) = \frac{1}{2} \cdot (f(x) - z)^2 + \lambda |p|$$ [6]

and hence, the Euler-Lagrange method, discussed in this thesis, can be applied to minimize the functional.

[6] ibid

References

Calder, Jeff. “The calculus of variations.” www-users.cse.umn.edu. April 29 2022. Web. May 1, 2022.

Christopher, Jason. “Using Calculus of Variations to Optimize Paths of Descent through Ski Race Courses.” https://dspace.mit.edu. June 2005. Web. March 14, 2022.

Cluckers, Raf. “Lipschitz continuity properties.” http://www.ub.edu. November 2008. Web. May 4, 2022.

Dawkins, Paul. “Calculus III - Higher Order Partial Derivatives.” http://tutorial.math.lamar.edu. Web. May 10, 2022.

GeoGebra. n.d. GeoGebra. https://www.geogebra.org. Web. February 14, 2022.

Gutermuth, Denise. “ Picard’s existence and uniqueness theorem.” https://ptolemy.berkeley.edu. November 2013. Web. May 2, 2022.

Liberzon, Daniel. “ Calculus of variations and Optimal Control Theory: A concise introduction.” Princeton University. Press. pp. 1-68.

Meng, Anders. “An Introduction to Variational Calculus in Machine Learning.” www2.imm.dtu.uk. February 2004. Web. April 16, 2022.

Miersemann, Erich. “Calculus of variations.” www.math.uni-leipzig.de. October 2012. Web. March 30, 2022.

O’Connor, J. J., & Robertson, E. F. “The brachistochrone problem.” mathshistory.st-andrews.ac.uk. February 2002. Web. May 27, 2022.

Simon, C. P., & Blume, L. E.” Mathematics for economists.” W. W. Norton & Company. Press. pp. 396-480.

Unknown. “Reasoning behind Second Partial Derivative Test (Article).www.khanacademy.org. Khanacademy. Web. March 20, 2022.