# Variational Principles

Lecture Notes

**Rashid M. Talha**

*School of Natural Sciences, NUST*
*February 16, 2023*
*Email: rmathsphys@gmail.com*

# Contents

# Abstract

Lecture notes for an introductory course on variational principles.
Currently a work-in-progess.

## Chapter 1

# Introduction

Suppose we want to find the total length of a curve $y = y(x)$ in $\mathbb{R}^2$, with $x \in [x_1, x_2]$, such that $y(x_1) = y_1$ and $y(x_2) = y_2$. The standard technique is to divide the curve into small line segments, each of length $dl$. Total length of $y$ is then

$$L[y] = \int dl. \tag{1}$$

Using Pythagoras' theorem, we can write $dl^2 = dx^2 + dy^2$. Moreover, $y = y(x)$ results in $dy = y'(x)dx$. The total length can, therefore, be expressed as

$$L[y] = \int_{x_1}^{x_2} \sqrt{1 + (y'(x))^2}dx. \tag{2}$$

The length $L$ will clearly depend on the given curve $y$, and the curve itself is a function of $x$. Therefore, $L$ can be treated as a function that maps a curve $y(x)$ to a real number $L[y]$. Such a map is called a functional, and it is the central object of study in the calculus of variations.

Now, suppose that we are not given any specific curve, rather we are asked to find the function $y(x)$ that satisfies $y(x_1) = y_1$ and $y(x_2) = y_2$ which has the minimum length $L[y]$. In other words, we are asked to minimise the functional $L[\cdot]$.

What do we *want* the answer to be? Based on our previous knowledge (or geometric analysis) we know that the curve of shortest length between any two distinct points on a plane is a straight line. However, it turns out that our current formulation of the problem does *not* give us this answer: the function

$$y = \begin{cases} y_1, & x = x_1 \\ y_2, & \text{otherwise} \end{cases} \tag{3}$$

satisfies the requirements of the problem, and has length less than that of a straight line segment joining $(x_1, y_1)$ and $(x_2, y_2)$.

So, we need to be improve our problem statement by adding more restrictions on $y$. We notice that the previous solution is discontinuous. So, we add the condition that $y$ needs to be at least continuous. However, this prevents us from using equation (2) to determine the length of the curve, since $y$ might not be differentiable. To remedy this, we require $y$ to be at least differentiable. (In fact, it turns out that we also need $y'(x)$ to be continuous, but we shall discuss such technicalities later in the course.)

Overall, we see that the problem of finding a minimum (or maximum) of a functional involves more than just the statement of the functional: it also requires us specify the boundary conditions and the *regularity* requirements, along with any other constraints. Similarly, there are several methods for finding the minimas and maximas that are applicable to specific functionals. In calculus of variations we approach this problem systematically by using the method of 'varying the dependent function around a critical point' in order to obtain the desired solution. The complete problem statement along with the solution strategy is called a *variational problem.*

In the following sections, we will study some of the basic terminology that is used in variational problems.

## 1.1 Regularity Classes

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function. We take $U$ to be an open set of $\mathbb{R}^n$ in order to ignore the complications due to boundary points.

**Definition 1.1.** A function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is called $k$-times differentiable if its $k$th derivative, $f^{(k)}$, exists.

**Definition 1.2** (Continuously differentiable)**.** A function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is called $k$-times continuously differentiable if it is $k$-times differentiable and the $k$th derivative, $f^{(k)}$, is continuous.

It is easy to check that if $f$ is $k$-times continuously differentiable then it is also $k-1$-times continuously differentiable, for $k \geq 1$. This follows from the fact that differentiability implies continuity, and that the existence of $f^{(k-1)}$ is a necessary condition for the existence of $f^{(k)}$.

The set of all $k$-times continuously differentiable functions on a domain $U$ is denoted by $C^k(U)$. When the domain is clear from the context (or irrelevant) then we simply write $C^k$.

$$C^k(U) = \{f : U \subseteq \mathbb{R}^n \to \mathbb{R} \mid f^{(k)} \text{ is continuous}\}. \tag{4}$$

$C^0$ denotes the set of all the continuous functions, and $C^\infty$ is the set of all such functions that can be differentiated infinitely many times.

**Example 1.3.** It is possible that a function $f$ is $k$-times differentiable, but that the $k$th derivative is *not* differentiable. In that case $f \in C^{k-1}$ only. For example, let

$$f(x) = \begin{cases} x^{k+2} \sin(\frac{1}{x}) & x \neq 0 \\ 0 & x = 0, \end{cases} \tag{5}$$

for $k = 0, 1, 2, \ldots$. Then, $f$ is $k+1$ times differentiable (for all $x \in \mathbb{R}$), but the $(k+1)$th derivative is not continuous (at $x = 0$). The regularity class of this function is, therefore, $C^k$; not $C^{k+1}$.

Similarly, we have the following observation

**Proposition 1.4.** *For all positive integers $k$, $C^{k-1} \subsetneqq C^k$, and $C^\infty \subsetneqq C^k$.*

*Proof.* TBC. □

## 1.2 Functionals

As briefly mentioned earlier, a functional $I$ is simply a function whose input is itself a function $f$, and output, denoted $I[f]$, is a real number. In other words, a functional maps functions to real numbers,

$$I : X \to \mathbb{R}, \qquad f \mapsto I[f] \in \mathbb{R}. \tag{6}$$

Here, $X$ is a set of functions of some suitable regularity. For example, equation (2) gives a functional whose domain can be $C^1([\alpha, \beta])$.

**Example 1.5.** All of the following are valid functionals:

- $I : C^0 \to \mathbb{R}$, with $I[f] = f(x_0)$ for some fixed $x_0 \in \mathbb{R}$.
- $I : C^0 \to \mathbb{R}$, with $I[f] = \int_0^1 f(x)^2 \, \mathrm{d}x$.
- $I : C^2 \to \mathbb{R}$, with $I[f] = f''(x_0)$ for some fixed $x_0 \in \mathbb{R}$.
- $I : C^0 \to \mathbb{R}$, with $I[f] = \int_0^1 f(x)^2 \sin(x^2) \, \mathrm{d}x / \int_0^1 f(x) \, \mathrm{d}x$, where the denominator is non-vanishing.

Calculus of variations primarily focuses on functionals of the form

$$I[y] = \int_{x_1}^{x_2} f\left(x, y(x), y'(x), \ldots, y^{(k)}(x)\right) \mathrm{d}x, \tag{7}$$

with $y \in C^r$ for some suitable positive integer $r$.

## Chapter 2

# Calculus in $\mathbb{R}^n$

## 2.1 Critical points and extreme values

Throughout this section let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function.

**Definition 2.1** (Global minimum)**.** The function $f$ has a global minimum at $a \in D$ if $f(a) \le f(x)$, for all $x \in D$.

**Definition 2.2** (Local minimum)**.** The function $f$ has a local minimum at $a \in D$ if there is an open set $U \subseteq D$, with $a \in U$, such that $f(a) \le f(x)$, for all $x \in U$.

**Definition 2.3** (Maximum)**.** The function $f$ has a local (global) maximum at $a \in D$, if $-f$ has a local (global) minimum at $a$.

We use the word *extremum* to mean either the maximum or the minimum.

Global extrema might not exist, and, in general, it is difficult to find them. Instead, we focus mainly on the local extrema.

**Proposition 2.4.** *If $f$ has a global minimum (maximum) at $a \in D$, then it has a local minimum (maximum) at $a$.*

**Definition 2.5** (Critical points)**.** Let $f$ be at least $C^1$. We say that $a \in D$ is a critical point of $f$, if $f'(a) = 0$.

Note that we have not defined critical points for functions that are not differentiable, or whose first derivative is not continuous. Similarly, we have ignored the complications due to boundary points.

**Theorem 2.6.** *Let $f$ be at least $C^1$. If $f$ has an extremum at $a \in D$, then $a$ is a critical point of $f$; that is $f'(a) = 0$.*

## 2.2 Convex functions

The converse of theorem (2.6) is not true in general. For example, a critical point can be a point of inflection or a saddle point, rather than being an extremum. If the function satisfies some additional properties then we can obtain more precise information about the extrema by studying only the critical points.

**Definition 2.7** (Convex function)**.** A function $f : D \to \mathbb{R}$ is called convex if $D$ is a convex set, and $\forall x, y \in D$ and $\forall t \in (0, 1)$,

$$(1-t)f(x) + tf(y) \geq f\left((1-t)x + ty\right). \tag{8}$$

If the inequality is strict for all $t \in (0, 1)$, then the function is called strictly convex.

Convex functions have particularly nice properties for optimisation problems. The following results discuss some of these features.

**Proposition 2.8.** *Suppose $f$ is convex, and that it has an extremum at $a \in D$. Then, $f$ has a global minimum at $a \in D$.*

**Proposition 2.9.** *Suppose $f \in C^1$ is convex, and that $f'(a) = 0$ for some $a \in U$. Then, $f'(x) \neq 0$ for all $x \neq a$. In other words, if there is a critical point then it is unique.*

EXAMPLE: F NOT CONVEX, LOCAL MIN IS NOT GLOBAL MIN.

## 2.3 Fundamental Theorem of Calculus of Variations

Before proceeding with the analysis of variational problems, we will study a very important result which is central in the derivation of the Euler-Lagrange equations.

Consider a continuous function $f : [x_1, x_2] \to \mathbb{R}$ such that

$$\int_{x_1}^{x_2} f(x)h(x)dx = 0 \tag{9}$$

for all the functions $h \in C^0([x_1, x_2])$. Does this imply that $f(x) = 0$ for all $x \in [x_1, x_2]$? By taking $h = f$, equation (9) becomes $\int_{x_1}^{x_2} f(x)^2 dx = 0$. This is the integral of a non-negative continuous function, and therefore, we obtain that $f(x) = 0$ for all $x \in [x_1, x_2]$ in this case. However, what happens when the 'test functions' $h(x)$ are restricted to a smaller set, such as $C^2$, or when there are boundary conditions on $h(x)$?

The next result, known as the fundamental theorem of calculus of variations (FTCV), presents this generalisation.

**Theorem 2.10** (FTCV version I)**.** *Let $x_1 < x_2$ and $k \in \mathbb{N}$. If $f : [x_1, x_2] \to \mathbb{R}$ is a continuous function such that*

$$\int_{x_1}^{x_2} f(x)h(x)dx = 0 \tag{10}$$

*for every function $h \in C^k([x_1, x_2])$ with $h(x_1) = h(x_2) = 0$, then $f(x) = 0$ for all $x \in [x_1, x_2]$.*

*Proof.* TBC. $\square$

**Exercise 2.1.** Prove that the function $h_0$ given above is $C^k$.

The test functions $h$ can also be restricted to the set $C^\infty$. The theorem still holds, and the proof remains exactly the same, but we need to chance $h_0$ so that it is infinitely differentiable. One one choice is

$$h_0(x) = \begin{cases} \exp(-1/(1-x^2)), & x \in (-1,1) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

This function can also be used in the previous proof.

There are several versions of FTCV, with slight modifications. The version presented in theorem (2.10) will be sufficient for us. For reference, a more general result is given below.

**Theorem 2.11** (FTCV version II)**.** *If $f : [x_1, x_2] \to \mathbb{R}$ is a continuous function such that*

$$\int_{x_1}^{x_2} f(x)h(x)dx = 0 \tag{12}$$

*for every function $h \in C^\infty([x_1, x_2])$ with $\operatorname{supp} h \subset (x_1, x_2)$, then $f(x) = 0$ for all $x \in [x_1, x_2]$.*

This version of FTCV implies theorem (2.10), because it is based on a strictly smaller set of assumptions.

# Chapter 3

# Euler-Lagrange equations

## 3.1 Some variational problems

The Brachistochrone, optical path, geodesics, classical mechanics.

## 3.2 Derivation of the Euler-Lagrange equations

Take $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$, with $x_1 < x_2$. Let $f \in C^2(\mathbb{R}^3)$ be a known function. The requirement that $f$ should be $C^2$ will become clearer later; having $\mathbb{R}^3$ as the domain of $f$ simply means that $f$ takes three real-valued inputs (which we treat as independent).

Consider the following variational problem:

**Problem.** Find an extremum[1] of the functional

$$I[y] = \int_{x_1}^{x_2} f(x, y(x), y'(x))dx \tag{13}$$

on the set of functions $y \in C^2([x_1, x_2])$ with $y(x_1) = y_1, y(x_2) = y_2$.

It is possible that such an extremum might not even exist. We shall explore this case later. Similarly, if an extremum exists then we don't known a priori if it is unique or of there are other extrema also. Moreover, currently we are not looking for a minimum (or a maximum), but only an extremum.

We begin solving this problem by supposing that a solution exists. For concreteness, assume that it is a minimum: that is, let $y(x)$ be a $C^2$ function satisfying the given boundary conditions that minimises the functional (13).

Now, we vary $y$ slightly while keeping its end points fixed. More precisely, we consider the function $\tilde{y}(x) = y(x) + hu(x)$, with $h \in \mathbb{R}$ sufficiently small, and $u \in C^2([x_1, x_2])$ with

[1] This will be some function $y(x)$ that satisfies the imposed regularity and boundary conditions.

$u(x_1) = u(x_2) = 0$. Then,

$$I[y] \le I[\tilde{y}] \equiv I[y + hu], \quad \forall h \text{ in some interval containing } 0, \tag{14}$$

since $y$ is a minimum of (13). It is easy to check that $\tilde{y} \in C^2([x_1, x_2])$ also, and that $\tilde{y}(x_1) = y_1$, $\tilde{y}(x_2) = y_2$. Observe that $\tilde{y}' = y' + hu'$.

Let $g_u : \mathbb{R} \to \mathbb{R}$ be defined as $g_u(h) = I[y+hu] \equiv I[\tilde{y}]$. Then, $g_u(0) = I[y]$. So, $g_u(0) \le g_u(h)$, for all sufficiently small $h$. In other words, $g_u$ has a minimum (i.e. an extremum) at $h = 0$. Therefore, $g_u'(0) = 0$.

So far we have seen that if $y$ minimises $I$ then $g_u'(0) = 0$.

Now,

$$g_u(h) = \int_{x_1}^{x_2} f\left(x, y(x) + hu(x), y'(x) + hu'(x)\right) dx = \int_{x_1}^{x_2} f\left(x, \tilde{y}(x), \tilde{y}'(x)\right) dx. \tag{15}$$

Therefore,

$$\begin{aligned} g_u'(h) &= \int_{x_1}^{x_2} \frac{d}{dh}\left(f\left(x, \tilde{y}(x), \tilde{y}'(x)\right)\right) dx & (16)\\ &= \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial x}\frac{dx}{dh} + \frac{\partial f}{\partial \tilde{y}}\frac{d\tilde{y}}{dh} + \frac{\partial f}{\partial \tilde{y}'}\frac{d\tilde{y}'}{dh}\right) dx & (17)\\ &= \int_{x_1}^{x_2} \left(u(x)\frac{\partial f}{\partial \tilde{y}} + u'(x)\frac{\partial f}{\partial \tilde{y}'}\right) dx & (18) \end{aligned}$$

Then,

$$g_u'(0) = \left.\frac{dg_u}{dh}\right|_{h=0} = \int_{x_1}^{x_2} \left(u(x)\frac{\partial f}{\partial y} + u'(x)\frac{\partial f}{\partial y'}\right) dx. \tag{19}$$

For technical clarity, note that the factors $\partial f/\partial y$ and $\partial f/\partial y'$ are short for

$$\frac{\partial f}{\partial y}(x, y, y') \quad \text{and} \quad \frac{\partial f}{\partial y'}(x, y, y'), \tag{20}$$

respectively. Integrating the second term in (19) by parts gives us

$$\begin{aligned} \int_{x_1}^{x_2} u'(x)\frac{\partial f}{\partial y'} dx &= u(x)\frac{\partial f}{\partial y'}\bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} u(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx & (21)\\ &= u(x_2)\frac{\partial f}{\partial y'} - u(x_1)\frac{\partial f}{\partial y'} - \int_{x_1}^{x_2} u(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx & (22)\\ &= -\int_{x_1}^{x_2} u(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx. & (23) \end{aligned}$$

Therefore, equation (19) becomes

$$g_u'(0) = \int_{x_1}^{x_2} \left(u(x)\frac{\partial f}{\partial y} - u(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right) dx = \int_{x_1}^{x_2} u(x)\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right) dx. \tag{24}$$

Consequently,

$$g_u'(0) = 0 \implies \int_{x_1}^{x_2} u(x) \left( \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) \right) dx \tag{25}$$

for all $u \in C^2$, with $u(x_1) = u(x_2) = 0$. Using the fundamental theorem of calculus of variations we conclude that

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) = 0. \tag{26}$$

At this point, we have established that if $y$ minimises $I$, then $y$ satisfies the equation (26).

Equation (26) is called the Euler-Lagrange equation associated to the variation problem with the functional (13).

**Remark.** If we had assumed that the extremum $y$ maximises $I$, then we would have still obtained $g_u(0) = 0$, leading to the same Euler-Lagrange equation. This can be seen most easily by replicating the above derivation with the assumption that $y$ maximises the given functional. This also suggests that if a function $y$ satisfies some Euler-Lagrange equations, then we can't in general conclude that it must be a minimum of the associated functional.

The proceeding analysis is formalised in the following theorem

**Theorem 3.1.** *THM*

In additional to the remark above, note that it is possible that a function $y$ might satisfy some Euler-Lagrange equations, but still not be an extremum of the associated functional. This is analogous to situation in real analysis where not every critical point is an extremum of the function. This observation motivates the following terminology.

**Definition 3.2.** EXTREMUM OF A FUNCTIONAL.

**Definition 3.3.** CRITICAL POINT OF A FUNCTIONAL.

EXAMPLE.

EXAMPLE.

## 3.3 Elementary analysis of the EL equations

Briefly introduce first integrals. We will return to these in a later chapter.

## 3.4 Higher derivatives

TBC

## 3.5 Multiple dependent variables

TBC

## 3.6 Multiple independent variables

TBC

## Appendix A

# Convexity and Legendre Transforms

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## 1.1 Convex Functions

## 1.2 Legendre Transforms

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## Appendix B

# Lagrange Multipliers

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## Appendix C

# More Examples of Variational Problems

Shortest path on R2, shortest path on S2, Path of least time, caternary: least energy, fermat's principle in optics, specific systems from CM and theoretical engineering, least area with fixed perimeter. Geodesics.

Advanced theoretical physics (String theory: Nambu-Gato action, Maxwell's equations from the action, comment about the lagrangian for the standard model).

# Bibliography


[1] R. Weinstock, Calculus of variations, with applications to physics and engineering, Dover Publications, 1974.

[2] F. Rindler, Calculus of variations, 1st Edition, Springer, 2018.

[3] I. M. Gelfand, S. V. Fomin, Calculus of variations, Prentice-Hall, 1963.

[4] L. D. Landau, E. M. Lifshitz, Mechanics, 2nd Edition, Pergamon Press, 1969.