2020 March 8

Scribed by Roh

hyunwoo@uchicago.edu

1 Calculus of Variation

The calculus of variations is a field of mathematics concerned with minimizing (or maximizing) functionals (that is, real-valued functions whose inputs are functions). It gives us precise analytical techniques to answer questions of the following type:

- Find the shortest path (i.e. geodesic) between two given points on a surface
- Find the curve between two given points in the plane that yields a surface of revolution of minimum area when revolved around a given axis
- Find the curve along which a bead will slide in the shortest time.

The calculus of variations is concerned with the problem of extremising "functionals." This problem is a generalisation of the problem of finding extrema of functions of several variables.In a sense to be made precise below, it is the problem of finding extrema of functions of an infinite number of variables. In fact, these variables will themselves be functions and we will be finding extrema of "functions of functions" or functional.

Just as an ordinary function takes a number as input and produces a number as output, a functional takes an entire function as input and produces a number. Many functionals are defined as **integrals over the input function**. The notation for a functional F with input function f is

$$F[f]$$

For example,

$$F[f] = \int_{-1}^{1} f(x)dx$$

If $f(x) = x^2$, then

$$\begin{aligned} F[x^2] &= \int_{-1}^{1} x^2 dx \\ &= \frac{2}{3} \end{aligned}$$

Just as regular function has a derivative with respect to its argument, a functional can have a functional derivative with respect to its input function. In a regular derivative, the idea is to change the independent variable (x for a function $f(x)$) a little bit (dx) and see how the function changes in response. A **functional derivative** changes the entire input function by a small amount $\delta f(x)$ and observes how the functional changes in response.

Obviously, there are an infinite number of ways we could change $f(x)$ in the functional; in the functional above, we might increase $f(x)$ a bit between -1 and 0 and decrease it a bit between 0 and $+1$, or we might increase or decrease it a bit over the entire range and so on. We clearly need something a bit more definite if we're get a consistent definition of a functional derivative.

1.1 Definition

The definition of a small amount, $\delta f(x)$, used in Lancaster & Blundell (Quantum Field Theory for the Gifted Amateur) is

$$\delta f(x) = \epsilon \delta(x - x_0)$$

where $\delta(x - x_0)$ is the Dirac delta function and ϵ is some small number, which means

$$\delta(x - x_0) = \begin{cases} 1 & x = x_0 \\ 0 & \text{otherwise} \end{cases}$$

The quantity x_0 is some value of x within the domain of $f(x)$. The idea is that the small change in $f(x)$ occurs at one point only (at $x = x_0$), which is why we are using Dirac delta function. With this definition, we can now define the functional derivative as

$$\frac{\delta F[f]}{\delta f(x_0)} \equiv \lim_{\epsilon \to 0} \frac{F\left[f(x) + \epsilon \delta(x - x_0)\right] - F\left[f(x)\right]}{\epsilon}$$

Note that δ is used in the notation $\frac{\delta F[f]}{\delta f(x_0)}$ for functional derivative, replacing $d \quad or \quad \partial$ in an ordinary derivative $\frac{df}{dx}$.

Example. With $F[f]$ defined above, we get

$$\begin{aligned} \frac{\delta F[f]}{\delta f(x_0)} &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\int_{-1}^{1} \left(f(x) + \epsilon \delta(x - x_0)\right) dx - \int_{-1}^{1} f(x) dx \right] \\ &= \int_{-1}^{1} \delta(x - x_0) dx \end{aligned}$$

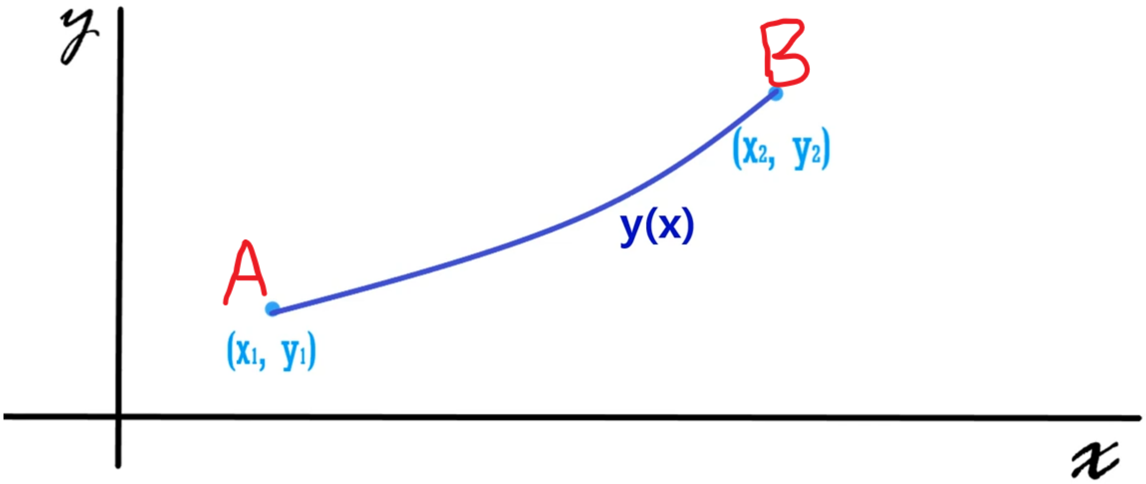
The value of the derivative depends on whether x_0 is within the range of integration, so we get

$$\frac{\delta F[f]}{\delta f(x_0)} = \begin{cases} 1 & \text{if } -1 < x_0 < 1 \\ 0 & \text{otherwise} \end{cases}$$

To see more example, you can find them in Here

1.2 A practical example of calculus of variations

To determine the shortest distance between given points $A(x_1, y_1)$ and $B(x_2, y_2)$ positioned in a two dimensional Euclidean space, calculus of variations will be applied to determine the functional form of the solution. Below is the illustration of the problem. We would like to find a function $y(x)$ that minimizes the line integral between two points.



The problem can be illustrated as minimizing the line-integral given by

$$I = \int_A^B 1 ds(x, y) \tag{1}$$

The above integral can be rewritten by a simple (and a bit heuristic) observations.

$$\begin{aligned} ds &= (dx^2 + dy^2)^{\frac{1}{2}} \\ &= dx \left[1 + \left(\frac{dy}{dx} \right)^2 \right]^{\frac{1}{2}} \end{aligned}$$

$$\frac{ds}{dx} = \left[1 + \left(\frac{dy}{dx} \right)^2 \right]^{\frac{1}{2}} \tag{2}$$

From this observation the line integral (length of the curve) can be written as

$$I(x, y, y') = \int_A^B \left[1 + \left(\frac{dy}{dx} \right)^2 \right]^{\frac{1}{2}} dx \tag{3}$$

The line integral given in the equation right above is also known as a functional, since it depends on x, y, y', where $y' = \frac{dy}{dx}$. Before performing the minimization, two important rules in variational calculus needs attention.

1.3 Rules of the game

- Functional derivative of a function with respect to itself is

$$\begin{aligned} \frac{\delta f(x)}{\delta f(y)} &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\phi(x) + \epsilon \cdot \delta(x - y) - \phi(x) \right] \\ &= \delta(x - y) \end{aligned}$$

This is the analogue of the ordinary derivative where

$$\frac{dy}{dx} = \begin{cases} 1 & y = x \\ 0 & \text{otherwise} \end{cases}$$

That is, if y and x are same independent variable, then the derivative is 1, but if they are different independent variables (that is, y is not a function of x), then the derivative is zero, since the two variables are, well, independent

- Communitative

$$\frac{\delta}{\delta f(x')} \frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x} \frac{\delta f(x)}{\delta f(x')}$$

Equivalently, with time derivative

$$\begin{aligned} \frac{\delta \dot{\phi}(t)}{\delta \phi(t_0)} &= \frac{d}{dt} \frac{\delta \phi(t)}{\delta \phi(t_0)} \\ &= \frac{d}{dt} \delta(t - t_0) \end{aligned}$$

- The $Chain\ Rule$ known from function differentiation applies.

The function $f(x)$ should be a smooth function, which usually is the case.

It is now possible to determine the type of functions which minimizes the functional given in equation (3). Before we proceed further, let's assume that we have a $y(x)$ and in addition to this we have small variation on that function which is described by $\epsilon \cdot \eta(x)$ as follows:

$$Y(x) = y(x) + \epsilon \cdot \eta(x)$$

We assume that $y(x)$ is the function that satisfies the minimization problem, which means that

$$\frac{\delta I(x, y, y')}{\delta y(x)} = 0 \tag{4}$$

or we can express it as

$$\frac{dI(x, y, y')}{d\epsilon} \mid_{\epsilon=0} = 0 \tag{5}$$

First, we will show how the calculus of variations works in as in equation (4). Then we will show exactly same thing in equation (5).

1.4 Equation(4)

To begin,

$$\frac{\delta I(x, y, y')}{\delta y(x_p)} = \frac{\delta}{\delta y(x_p)} \int_A^B \left[1 + \left(\frac{dy}{dx} \right)^2 \right]^{\frac{1}{2}} dx$$

where we now define $y_p \equiv y(x_p)$ to simplify the notation. Using " the rules of the game", it is possible to differentiate the expression inside the integral to give

$$
\begin{aligned}
\frac{\delta I(x,y,y')}{\delta y(x_p)} &= \frac{\delta}{\delta y(x')}\int_A^B \left[1+\left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} dx \\
&= \frac{\delta}{\delta y_p}\int_A^B \left[1+\left(y'\right)^2\right]^{\frac{1}{2}} dx \qquad \because \text{definition} \\
&= \int_A^B \left(\frac{1}{2}\cdot\left[1+\left(y'\right)^2\right]^{-\frac{1}{2}}\cdot 2y'\cdot\left(\frac{\delta y'}{\delta y_p}\right)\right) dx \\
&\text{we multiply } \left(\frac{\delta y'}{\delta y_p}\right) \text{ because we differentiate on } y_p \text{ not } x \\
&= \int_A^B \left(\frac{1}{2}\cdot\left[1+\left(y'\right)^2\right]^{-\frac{1}{2}}\cdot 2y'\cdot\left(\frac{\delta}{\delta y_p}\cdot\frac{dy}{dx}\right)\right) dx \\
&= \int_A^B \left(\frac{y'}{\left[1+\left(y'\right)^2\right]^{\frac{1}{2}}}\cdot\frac{d}{dx}\frac{\delta y(x)}{\delta y(x_p)}\right) dx \qquad \because \text{Communitative} \\
&= \int_A^B \left(\underbrace{\frac{y'}{\left[1+\left(y'\right)^2\right]^{\frac{1}{2}}}}_{g}\cdot\underbrace{\frac{d}{dx}\delta(x-x_p)}_{f}\right) dx
\end{aligned}
$$

The chain rule in integration:

$$\int f(x)g(x)dx = F(x)g(x) - \int F(x)g'(x)dx$$

can be applied to the expression above, which results in the following equations

$$
\begin{aligned}
\frac{\delta I(x,y,y')}{\delta y(x_p)} &= \left[\frac{y'}{\left[1+\left(y'\right)^2\right]^{\frac{1}{2}}}\delta(x-x_p)\right]_A^B - \int_A^B \delta(x-x_p)\frac{d}{dx}\left(\frac{y'}{\left[1+\left(y'\right)^2\right]^{\frac{1}{2}}}\right) dx \\
&= \left[\frac{y'}{\left[1+\left(y'\right)^2\right]^{\frac{1}{2}}}\delta(x-x_p)\right]_A^B - \int_A^B \delta(x-x_p)\frac{y''}{\left[1+\left(y'\right)^2\right]^{\frac{3}{2}}} dx
\end{aligned}
$$

Now assuming that on the boundaries, A and B, the first part of the expression given in the equation right above disappears, so

$$\delta(A-x_p) = \delta(B-x_p) = 0$$

then the expression simplifies to (assuming that $\int_A^B \delta(x-x_p)dx = 1$)

$$\frac{\delta I(x,y,y')}{\delta y(x_p)} = \frac{y_p''}{\left[1+\left(y_p'\right)^2\right]^{\frac{3}{2}}}$$

And set this to zero and determine the function form of $y(x_p)$:

$$y_p'' = 0$$

Solution

A solution to the differential equation $y_p'' = 0$ is $y_p(x) = ax + b$; the family of straight lines, which intuitively makes sense. We already know that the straight line is the answer. However in cases where the space is not Euclidean,the functions which minimized the integral expression or functional may not be that easy to determine. In the above case, to guarantee that the found solution is a minimum the functional have to be differentiated once more with respect to $y_p(x)$ to make sure that the found solution is a minimum.

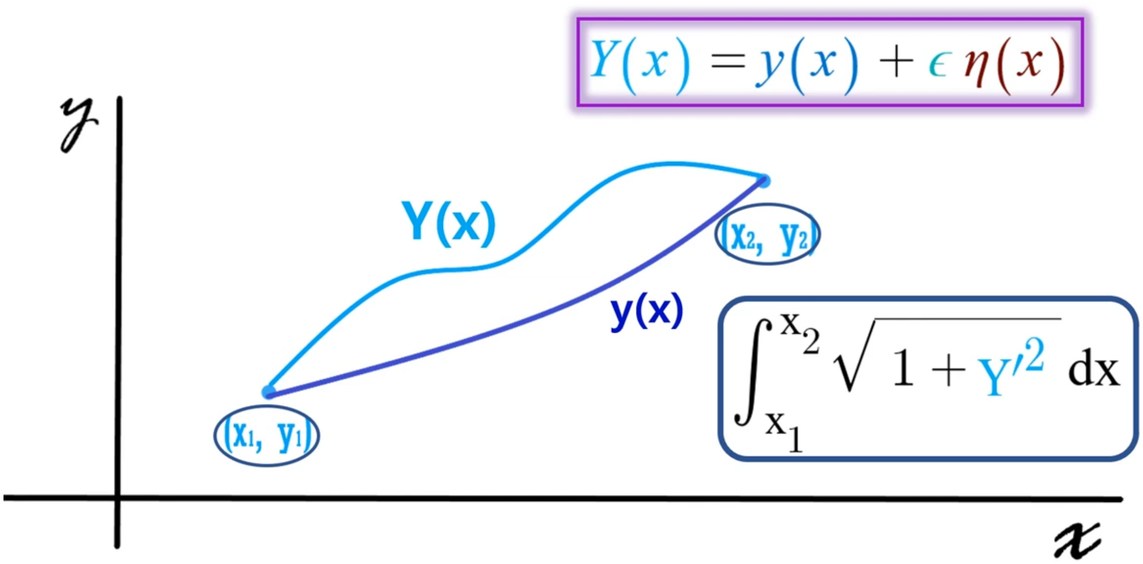
I will continue to talk how this is related to some techniques in machine learning.

1.5 Equation (5)

To begin with

$$\frac{dI(x,y,y')}{d\epsilon}\,|_{\epsilon=0}= 0$$

Below is the picture of our setting.



To continue,

$$\begin{aligned}\frac{dI(x,y,y')}{d\epsilon} &= \int_A^B \left(\frac{1}{2}\cdot\left[1+(Y')^2\right]^{-\frac{1}{2}}\cdot 2Y'\cdot\left(\frac{dY'}{d\epsilon}\right)\right)dx\\ &= \int_A^B \left(\frac{Y'}{\left[1+(Y')^2\right]^{\frac{1}{2}}}\cdot\eta'(x)\right)dx\end{aligned}$$

evaluate at $\epsilon = 0$ then Y' is replaced by y'

$$\Longrightarrow \int_A^B \left(\frac{y'\eta'(x)}{\left[1+(y')^2\right]^{\frac{1}{2}}}\right)dx = 0$$

Then we do chain rule to get,

$$
\begin{aligned}
\frac{dI(x,y,y')}{d\epsilon}\ |_{\epsilon=0} &= \left[\frac{y'}{\sqrt{1+(y')^2}}\boldsymbol{\eta(x)}\right]_A^B - \int_A^B \boldsymbol{\eta(x)}\frac{d}{dx}\left(\frac{y'}{\left[1+(y')^2\right]^{\frac{1}{2}}}\right)dx \\
&= 0 \qquad - \qquad \int_A^B \boldsymbol{\eta(x)}\frac{d}{dx}\left(\frac{y'}{\left[1+(y')^2\right]^{\frac{1}{2}}}\right)dx \\
&= -\int_A^B \boldsymbol{\eta(x)}\frac{d}{dx}\left(\frac{y'}{\left[1+(y')^2\right]^{\frac{1}{2}}}\right)dx
\end{aligned}
$$

For this to be zero with arbitrary function $\eta(x)$, this means that

$$
\begin{aligned}
&\frac{d}{dx}\left(\frac{y'}{\left[1+(y')^2\right]^{\frac{1}{2}}}\right) = 0 \\
\iff &\frac{y'}{\left[1+(y')^2\right]^{\frac{1}{2}}} = C \\
\iff &(y')^2 = C^2\left(1+(y')^2\right) \\
\iff &(y')^2 = \frac{C^2}{1-C^2} \\
\iff &y' = C \\
\iff &y = ax+b \qquad \text{(straight line)}
\end{aligned}
$$

1.6 Euler-Lagrange Equation

In general form, we can think of functional I

$$I = \int_A^B F(x, Y, Y')dx$$

Similarly, we take a derivative w.r.t ϵ to get

$$\frac{dI}{d\epsilon} = \int_A^B \left(\frac{\partial F}{\partial \epsilon}\right)dx$$

We can take the derivative direct on the integrand because the integration is over x which is independent of ϵ. To continue by chain rule

$$
\begin{aligned}
&\int_A^B \left[\frac{\partial F}{\partial Y}\frac{\partial Y}{\partial \epsilon} + \frac{\partial F}{\partial Y'}\frac{\partial Y'}{\partial \epsilon}\right]dx \\
&= \int_A^B \left[\frac{\partial F}{\partial Y}\eta(x) + \frac{\partial F}{\partial Y'}\eta'(x)\right]dx \\
&\text{evaluate at } \epsilon = 0 \text{ then } Y' \text{ is replaced by } y' \\
\implies &= 0
\end{aligned}
$$

Using chain rule on the second term above to get

$$\frac{\partial F}{\partial Y'}\eta'(x) = \underbrace{\left[\frac{\partial F}{\partial y'}\boldsymbol{\eta(x)}\right]_A^B}_{=0} - \int_A^B \boldsymbol{\eta(x)}\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)dx$$

Therefore,

$$\begin{aligned} &\int_A^B \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right]\eta(x)dx = 0 \\ \Longleftrightarrow\ &\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \end{aligned}$$

This is the Euler-Lagrange Equation.

Reference:

1. https://physicspages.com/pdf/Lancaster%20QFT/Lancaster%20Problems%2001.02.pdf

2. https://physicspages.com/pdf/Lancaster%20QFT/Lancaster%20Problems%2001.03-04.pdf

3. https://www.youtube.com/watch?v=08vJyA-XD3Q

4. https://www.youtube.com/watch?v=OcRB6omfy9c (Korean, similar to reference 3)