Calculus of Variations

Akash Garg

Version 1.1 - February 19, 2023

1 - Functions vs. Functionals

The fundamental concept behind variational calculus is the concept of a functional. A *functional* assigns a real number $\Re$ to each function belonging to a class, it's a function of functions. Typical examples of a functional are the length of a curve, center of mass of a curve, a path between two points in space, etc. In machine learning, the cost function of a neural network is also considered a functional, where it optimizes/minimizes a given $\mathbf{f}$, a function represented by a neural network. As an illustrative example of the calculus of variations, we will consider the length of a curve:

The length of a curve in the interval $[a, b]$ can be written as:

$$l = \int_a^b \sqrt{1 + y'(x)^2} \mathrm{d}x$$

A functional is usually of the form:

$$v = \int_a^b F(x, y(x), y'(x)) \mathrm{d}x,$$

where F is itself a function that depends on the free variable x and the functions $y(x)$ and $y'(x)$.

One can consider,

$$v = \int_a^b F(x, y(x), y'(x)) \mathrm{d}x \qquad y(a) = A, y(b) = B$$

and use the points $a = x_0, x_1, \ldots, x_n, x_{n+1} = b$ to divide the interval $[a, b]$ into $n+1$ equal parts and replace $y = y(x)$ such that we have the points:

$$(x_0, A), (x_1, y(x_1)), \ldots, (x_n, y(x_n)), (x_{n+1}, B).$$

Then we can approximate v in the following manner:

$$v(y, \ldots, y_n) = \sum_i^{n+1} F(x_i, y_i, \frac{y_i - y_{i-1}}{h})h, \qquad y_i = y(x_i), h = x_i - x_{i-1}$$

In solving variational problems Euler made extensive use of finite differences and then obtained exact solutions by taking the limit as $h \to 0$. In this sense a functional v can be regarded as a function of an ∞ variables.

Three major problems influenced the development of variational calculus: 1. Brachistochrone: developed by Bernoulli, it is a problem of finding the path of quickest descent of a particle between two points A and B under the influence of external forces like gravity. Turns out that the curve of quickest descent is a cycloid. 2. Geodesics: define a line of minimum length lying on the surface $\phi(x, y, z) = 0$ and joining two given points on a surface. This is an example of finding a constrained minima, where the constraint is that the line must lie on the surface. This was originally solved by Bernoulli, but a generlized method was given by Euler & Lagrange. 3. Isoperimetric: is a closed curve of fixed length that encircles a maximal area. Turns out that the solution to this problem is a cirlce. This is another example of contraint maxima, where the length of the curve $l = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} dt$ is kept constant while maximizing it's encompassing area.

Variational calculus is the only developed branch of the "calculus of functionals" concerning itself with the maxima and minima of a functional.

Finding the extremum points of a functional is very much like the finding the extremum points of a regular function. Here are a few similarities between the two:

Ordinary functions	Functionals
$z = f(x)$	$v = v(y(x))$
$\Delta x = x - x_i, dx = \Delta x$	$\delta y = y(x) - y_i(x)$ Here assume that y is running through certain class of functions.
Continuous if small variations in x leads to small variations in $f(x)$	Continuous if small variations in $y(x)$ leads to small variations in $v(y(x))$. Intuitively, one can think of a small variation in $y(x)$ as slightly "wiggling" the curve $y(x)$ to get an alternate curve.

Let's take a small departure to discuss what it means for a curve $y(x)$ to be close to another curve $y_i(x)$. A curve $y(x)$ and can be close to another curve $y_i(x)$ by comparing differences in positions, tangents, or higher order properties. The distance between two curves can be measured to zero'th order by using:

$$\max(|y(x) - y_i(x)|).$$

The above only looks at positional data. When dealing with a class of curves, we can measure distances in higher orders. For example, the distance between two curves up to first order is:

$$\max(|y(x) - y_i(x)|) + \max(|y'(x) - y_i'(x)|).$$

Similarly we can extend this out to support a measure of closeness to any order:

$$\sum_{j=0}^{k} \max(|y^{(j)}(x) - y_i^{(j)}(x)|)$$

Using the above defintion, we can redefine the relationship between variations of functions and functionals more formally.

Ordinary functions	Functionals
$\lvert f(x) - f(x_0)\rvert < \epsilon$ when $\lvert x - x_0\rvert < \delta$	$\lvert v(y(x)) - v(y_0(x))\rvert < \epsilon$ when $\begin{aligned} &\lvert y(x) - y_0(x)\rvert && < \delta \\ &\lvert y'(x) - y_0'(x)\rvert && < \delta \\ &\lvert y''(x) - y_0''(x)\rvert && < \delta \\ &\vdots \\ &\lvert y^{(k)}(x) - y_0^{(k)}(x)\rvert && < \delta \end{aligned}$
$f(cx) = cf(x)$ $l(x_1 + x_2) = l(x_1) + l(x_2)$	$v(c(y(x))) = cv(y(x))$ $v(y_1(x) + y_2(x)) = v(y_1(x)) + v(y_2(x))$
$\Delta f = f(x + \Delta x) - f(x)$. Doing a taylor expansion, we get two terms: $\Delta f = A(x)\Delta x + B(x, \Delta x)\Delta x$ where, $B \to 0$ as $\Delta x \to 0$. Here $B(x, \Delta x)$ are the higher order terms and $A(x)$ is the linear part of the differential, and thus we get: $df = f'(x)\Delta x$	$\Delta v = v(y(x) + \delta y) - v(y(x))$. Doing the taylor expansion we get two terms: $\Delta v = L(y(x), \delta y) + B(y(x), \delta y)\max(\lvert\delta y\rvert)$, where $B \to 0$ as $\max(\delta y) \to 0$. Here $B(y(x), \delta y)$ are the higher order terms and $L(y, \delta y)$ is the linear part, and thus we get: $\delta v = L(y(x), \delta y)$.

Let's take that last fact and define differentials in another way. We define a paramater α and use it as follows:

$$f(x + \alpha\Delta x),$$

when $\alpha = 1$ we obtain the increased value for f, and when $\alpha = 0$ we obtain the original value for f. Then we can differentiate with respect to α and get:

$$\frac{\partial}{\partial \alpha} f(x + \alpha\Delta x) = f'(x + \alpha\Delta x)\Delta x|_{\alpha=0} = f'(x)\Delta x = df(x).$$

Similarly, we can use it to find the differentail fo multivariate functions:

$$f(x_1 + \alpha\Delta x, \ldots, x_n + \alpha\Delta x)$$

$$\frac{\partial}{\partial \Delta x} f(x_1 + \alpha\Delta x, \ldots, x_n + \alpha\Delta x) = \nabla f \cdot [\Delta x_1, \ldots, \Delta x_n]^{\mathrm{T}} = \sum_i^n \frac{\partial f}{\partial x_i}\Delta x_i = df$$

Note that evaluating α at zero gives us what we are looking for, the differential of f, df.

Let's carry over the same methodology to evaluate the differential of a functional. First we know:

$$\Delta v = v(y(x) + \alpha\delta y) - v(y(x)) = L(x, \alpha\delta y) + B(y, \alpha\delta y)|\alpha| \max(|\delta y|)$$

Here, we know that L is the linear part of the difference and B is the higher-order part; this comes directly from taking a taylor expansion of Δv. To find the differential of v, we evaluat the derivative of α at zero.

$$\begin{aligned} \lim_{\alpha \to 0} \frac{\Delta v}{\alpha} &= \lim_{\alpha \to 0} \frac{L(y, \alpha\delta y) + B(y, \alpha\delta y)|\alpha| \max(|\delta y|)}{\alpha} \\ &= \lim_{\alpha \to 0} \frac{L(y, \alpha\delta y)}{\alpha} + \underbrace{\lim_{\alpha \to 0} \frac{B\cdots}{\alpha}}_{0} \end{aligned}$$

$$L(y, \alpha\delta y) = \alpha L(y, \delta y),$$

by linearity, then,

$$\lim_{\alpha \to 0} \frac{L(y, \alpha\delta y)}{\alpha} = \lim_{\alpha \to 0} \frac{\alpha L(y, \delta y)}{\alpha} = L(y, \delta y) = \delta v$$

2 - Fundamental Theorem of Variational Calculus

We know from calculus of functions that $x = x_0$ extremize the function $f(x)$ when $df = 0$. When $df = 0$, we know that the function $f(x)$ is extremized but it tells us nothing about a minimum/maximum. If small variations in x increase $f(x)$, then x minimizes it, otherwise if small variations in x decrease $f(x)$, then x maximimizes it.

Similarly, the fundamental theorem of the calculus of functionals states that a curve $y(x) = y_0(x)$ extremizes the functional $v(y(x))$ when $\delta v = 0$ along $y_0(x)$. In other words when small "wiggles" of curve cause no change in v. Again this condition tells us that $y_0(x)$ is an extremizing curve, but it doesn't tell us if it is minimizing/maximizing. We can do the same test to see if it is minimizing/maximing v, by checking if variations in $y_0(x)$ increase or decrease the value of v.

2.1 - Strong vs. Weak Extrema

Recall that there are various different orders of "closeness" that can be measured for a functional. Since taking variations of a curve involves some notion of measuring "closeness", i.e., we want to find curves that deviate slightly from the current curve, different forms of exterma's can be realized for functionals. If $y = y_0(x)$ extremizes $v(y(x))$ in all classes of curves such that $|y(x) - y_0(x)|$ is small, i.e., in the class of all neighboring curves of $y = y_0(x)$ in the sense of closeness of order zero, then such extrema of v is *strong*.

Similarly, if $y = y_0(x)$ extremizes $v(y(x))$ only for a class of curves in the first-order sense then such an extrema is *weak*. Note that a strong extrema is also a weak extrema, but not the other way around.

At first glance this can be confusing, why does stricter form of "closeness"cause a weaker extrema? The answer is simple, the stricter the rule for "closeness", the less curves there exist in that class and as a result the extrema is weaker. For example, all zero-order close curves are also first-order close curves, but not the other way around. In other words, zero-order close curves are a super-set of all higher-order closeness.

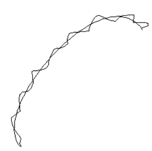

Figure 1: Zero-Order Approximation

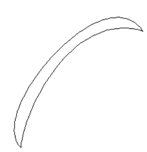

Figure 2: First-Order Approximation

3 - The Euler-Lagrange Equation

We are now at a point where we can use hte machinery laid out for us to derive the famous Euler-Lagrange equations, a PDE that is useful in solving many variational problems. The Euler-Lagrange equations can be derived when we wish to extremize a functional of the form:

$$v(y(x)) = \int_a^b F(x, y(x), y'(x))\mathrm{d}x \tag{1}$$

Before we continue we take one small departure. Assume that $y = y(x)$ and any admissible curve $y = y^*(x)$ neighboring $y(x)$. Then we can set $y(x, \alpha) = y(x) + \alpha(y^*(x) - y(x))$, when $\alpha = 0$ we get $y(x)$ and when $\alpha = 1$ we get $\delta y = y^*(x) - y(x)$. Note that δy is a function of x and is differentiable. This leads to an interesting observation:

$$\begin{aligned} (\delta y)' &= (y^*(x) - y(x))' = y^*(x)' - y(x)' = \delta y' \\ (\delta y)'' &= (y^*(x) - y(x))'' = y^*(x)'' - y(x)'' = \delta y'' \\ (\delta y)^{(k)} &= (y^*(x) - y(x))^{(k)} = y^*(x)^{(k)} - y(x)^{(k)} = \delta y^{(k)} \end{aligned} \tag{2}$$

This leads to an interesting fact that will be useful to to us later: *derivative of the variation = variation of the derivative*.

We wish to extremize v in Eq. 1 such that $\delta v(y_0(x)) = 0$ and $y_0(a) = A, y_0(b) = B$. In order to do this we take the variation of v and set it to zero.

$$\begin{aligned} \Delta v = v(y(x) + \delta y) - v(y(x)) &= \int_a^b F(x, y + \delta y, y' - \delta y')\mathrm{d}x - \int_a^b F(x, y, y')\mathrm{d}x \\ &= \int_a^b F(x, y + \delta y, y' + \delta y') - F(x, y, y')\mathrm{d}x \end{aligned}$$

Using Taylor expansion, we can approximate $F(x, y + \delta y, y' + \delta y')$ as

$$F(x, y + \delta y, y' + \delta y') = \underbrace{F(x, y, y') + F_y(x, y, y')\delta y + F_{y'}(x, y, y')\delta y'}_{\text{Pricipal linear part}} + \underbrace{\cdots}_{\text{higher order terms}}$$

We can ignore hte higher order terms since for small variations of δy we only care about the principal linear component. We can plug this in above and we get the following relation:

$$\begin{aligned} \Delta v = v(y(x) + \delta y) - v(y(x)) &= \int_a^b F_y(x, y, y')\delta y + F_{y'}(x, y, y')\delta y'\mathrm{d}x \\ &= \int_a^b F_y(x, y, y')\delta y\mathrm{d}x + \int_a^b F_{y'}(x, y, y')\delta y'\mathrm{d}x = \delta v \end{aligned}$$

We wish to solve for $y(x)$ such that $\delta v = 0$. We can solve the above using integration by parts. Recall integration by parts:

$$\int u dw = uw - \int w du$$

For us, $u = F_{y'}$ and $w = \delta y$. We first solve the second part of the integral involving the $F_{y'}$ term using integration by parts:

$$\int_a^b F_{y'}(x, y, y')\delta y' = F_{y'}\delta y|_a^b - \int_a^b \frac{d}{dx}F_{y'}\delta y$$

Since we want curves that pass through the endpoints, there is no variations evaluated at the end points and as a result the term δy evaluated at a and b is zero. As a result:

$$\int_a^b F_{y'}(x,y,y')\delta y' = -\int_a^b \frac{d}{dx}F_{y'}\delta y \tag{3}$$

Plugging this back in, we get

$$\delta v = \int_a^b (F_y - \frac{d}{dx}F_{y'})\delta y \mathrm{d}x = 0 \tag{4}$$

Before we continue, we need to make use of the fundamental lemma of variational calculus.

3.1 - Fundamental Lemma of Variational Calculus

If we have a functional of the form:

$$\int_a^b \phi(x)f(x)\mathrm{d}x = 0,$$

then, for any arbitrary function $f(x), phi(x) \equiv 0$ for $a \leq x \leq b$. Suppose that $x = x^*$ such that $a \leq x^* \leq b$ and $\phi(x^*) \neq 0$. If $\phi(x)$ is continuous, then there is a neighborhood $x_0^* \leq x \leq x_1^*$ at point x^* throughout which $\phi(x)$ has a constant sign.

If we choose $f(x)$ so that it has a constant sign in that neighborhood and zero elsewhere, we get:

$$\int_a^b \phi(x)f(x)\mathrm{d}x = \int_{x_0^*}^{x_1^*} \phi(x)f(x)\mathrm{d}x \neq 0,$$

A contradiction, and hence $\boldsymbol{\phi(\mathbf{x}) \equiv 0}$

3.2 - Euler-Lagrange Equation

Now we can apply the result from §3.1 to solve Equation Eq. 4. Here we notice that δy is the arbitrary function and hence we can say that:

$$\delta v = \int_a^b (F_y - \frac{d}{dx}F_{y'})\delta y \mathrm{d}x = 0 \Rightarrow F_y - \frac{d}{dx}F_{y'} = 0 \tag{5}$$

Solving the following PDE gives us the function that extremizes v:

$$F_y - \frac{d}{dx}F_{y'} = F_y - F_{xy'} - F_{yy'}y' - Fy'y'y'' = 0 \tag{6}$$

3.3 - Shortest Path Between Two Points

A simple example to showcase the result in Eq. 6, let's do a simple problem, finding the shortest path between any two points. The functional that captures the lenght of the path between two points a and b:

$$\int_a^b \sqrt{1 + y'^2}dx$$

Plugging into the euler equation, we get:

$$\begin{aligned} F_y &= 0 \\ F_{xy'} &= 0 \\ F_{y'} &= (1+y'^2)^{-\frac{1}{2}}y' \\ F_{yy'} &= 0 \\ F_{y'y'} &= (1+y'^2)^{-\frac{1}{2}} - y'^2(1+y'^2)^{-\frac{3}{2}} \end{aligned}$$

The only equation that survives is the last term

$$F_{y'y'}y'' = \left((1+y'^2)^{-\frac{1}{2}} - y'^2(1+y'^2)^{-\frac{3}{2}}\right)y'' = 0$$

Simplifying this out, we get:

$$\left(\frac{1}{\sqrt{1+y'^2}} - \frac{y'^2}{(1+y'^2)^{\frac{3}{2}}}\right)\cdot y'' = 0$$

Simplifying the above further, we get:

$$\frac{y''}{(1+y'^2)^{\frac{3}{2}}} = 0$$

Recall that the above is just the curvature equation and it says that all curves that the path that minimizes the path between two points a and b is given by a curve with zero curvature, i.e., a straight line. An elementary fact that we are all familiar with is proven here using the calculus of variations.

Revision	**Date**	**Description**
1.0	April 15, 2012	Creation
1.1	February 19, 2023	Reformatting