Calculus of Variations*

(Com S 477/577 Notes)

Yan-Bin Jia

Nov 17, 2020

1 Introduction

A *functional* assigns a real number to each function (or curve) in some class. One might say that a functional is a function of another function (or curve). Functionals have an important role in many problems arising in analysis, optimization, mechanics, geometry, etc.

Example 1. Consider all possible paths joining two points $\boldsymbol{p}$ and $\boldsymbol{q}$ in the plane, and a particle moving along any of these paths with velocity $\boldsymbol{v}(x, y)$ at the point $(x.y)$. Then we obtain a functional that assigns to each path the time of traversal by the particle.

Example 2. Let $y(x)$ be an arbitrary continuously differentiable function defined on the interval $[a, b]$. Then the integral

$$J[y] = \int_a^b y'^2(x)\, dx$$

is a functional on the set of all such functions $y(x)$.

Example 3. Let $F(x, y, z)$ be a continuous function of three variables. Then the integral

$$J[y] = \int_a^b F(x, y(x), y'(x))\, dx,$$

where $y(x)$ ranges over the set of all continuously differentiable functions defined over the interval $[a, b]$, is a functional. Different choices of the function $F(x, y, z)$ give different functionals. For example, if

$$F(x, y, z) = \sqrt{1 + z^2},$$

then $J[y]$ is the length of the curve $y(x)$. If

$$F(x, y, z) = z^2,$$

$J[y]$ reduces to the case considered in the previous example.

Instances of problems involving functionals were considered more than three hundred years ago, and the first important results are due to Euler (1707-1783). Though still not having methods

*The material is adapted from the book *Calculus of Variations* by I. M. Gelfand and S. V. Fomin, Prentice Hall Inc., 1963; Dover, 2000.

comparable to those of classical analysis in generality, the "calculus of functionals" has had the most development in finding the extrema of functionals, a branch referred to as the "calculus of variations." Listed below are a few typical *variational problems*.

1. Find the shortest plane curve connecting two points $\boldsymbol{p}$ and $\boldsymbol{q}$, that is, find the curve $y = y(x)$ for which the functional
$$\int_a^b \sqrt{1 + y'^2}\, dx$$
achieves its minimum. The shortest curve turns out to be the line segment $\overline{\boldsymbol{pq}}$.

2. Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be two points in the vertical plane[1]. Consider a particle sliding under gravity along some path (curve) joining $\boldsymbol{p}$ and $\boldsymbol{q}$. The time it takes the particle to reach $\boldsymbol{q}$ from $\boldsymbol{p}$ depends on the curve, and hence is a functional. The curve that results in the least time is called the *brachistochrone*. The brachistochrone problem was posed by John Bernoulli in 1696, and solved by himself, James Bernoulli, Newton, and L'Hospital. The brachistochrone turns out to be a cycloid, as we will discuss later.

3. The *isoperimetric* problem was solved by Euler: Among all closed plane curves of a given length, find the one that encloses the greatest area. The answer turns out, not surprisingly, to be a circle.

All of the above problems can be written in the form
$$\int_a^b F(x, y, y')\, dx.$$
Such a function has a "localization property" that the value of the functional equals the sum of its values over segments generated by dividing the curve $y = y(x)$. Below is a functional example that does not have this property:
$$\frac{\int_a^b x\sqrt{1 + y'^2}\, dx}{\int_a^b \sqrt{1 + y'^2}\, dx}.$$

It is important to see how problems in calculus of variations are related to those of classical analysis, especially, to the study of functions of n variables. Consider a functional of the form
$$J[y] = \int_a^b F(x, y, y')\, dx, \qquad y(a) = y_a, \quad y(b) = y_b.$$
To see the connection, we divide the interval $[a, b]$ into $n + 1$ equal parts:
$$a = x_0, x_1, \ldots, x_n, x_{n+1} = b.$$
Then, replace the curve $y = y(x)$ by the polygonal line with vertices
$$(x_0, y_a), (x_1, y(x_1), \ldots, (x_n, y(x_n)), (x_{n+1}, y_b).$$
Essentially, we are approximating the functional $J[y]$ by the sum
$$\tilde{J}(y_1, \ldots, y_n) = \sum_{i=1}^{n+1} F\left(x_i, y_i, \frac{y_i - y_{i-1}}{h}\right) h,$$

[1]The two points are assumed to have distinct x-coordinate values.

where $y_i = y(x_i)$ and $h = \frac{b-a}{n+1}$. Thus, we can regard the variational problem as that of finding the extrema of the function $J(y_1, \ldots, y_n)$, with $n \to \infty$. In this sense, functionals can be regarded as "functions of infinitely many variables", and the calculus of variations as the corresponding analog of differential calculus.

2 The Variation of a Functional

In a variational problem, we treat each function belonging to some class as a point in some space referred to as the *function space*. For instance, if we are dealing with a functional of the form

$$\int_a^b F(x, y, y')\, dx,$$

the space includes all functions with a continuous first derivative. In the case of a functional

$$\int_a^b F(x, y, y', y'')\, dx,$$

the appropriate function space is the set of all functions whose second derivatives are continuous. For a variational problem, we consider the linear space of feasible functions.

We need to measure the "closeness" between two elements in a function space. This is done by introducing the concept of the *norm* of a function.[2] Let the space $\mathcal{D}$, or more precisely $\mathcal{D}(a,b)$, consist of all continuous functions defined on an interval $[a,b]$. Addition and multiplication by real numbers are in the usual sense. The norm is defined as

$$\|y\|_0 = \max_{a \le x \le b} |y(x)|.$$

As illustrated in Figure 1, the distance between the function $y(x)$ and an "optimal" function $y^*(x)$ does not exceed ϵ if the graph of $y(x)$ lies inside a band of width 2ϵ centered at the graph of $y^*(x)$.

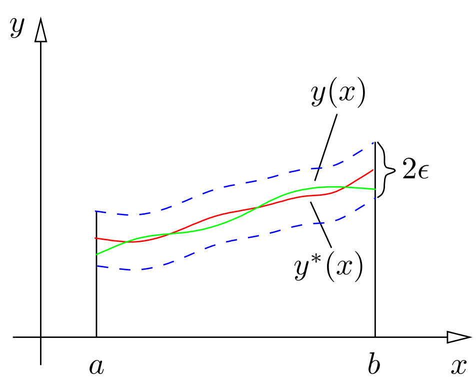

Figure 1: Distance between two functions is the maximum absolute difference of their values at any point.

Similarly, the space $\mathcal{D}_n$, or more precisely $\mathcal{D}_n(a,b)$, consists of all functions defined on $[a,b]$ that are n times continuously differentiable. The norm now is defined as

$$\|y\|_n = \sum_{i=0}^{n} \max_{a \le x \le b} \left| y^{(i)}(x) \right|.$$

A functional $J[y]$ is said to be *continuous* at a point (i.e., a function) y_0 in a normed function space $\mathcal{F}$ if for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|J[y] - J[y_0]| < \epsilon,$$

for all $\|y - y_0\| < \delta$. A continuous functional J is said to be *linear* if

[2] We have indeed used the concept of the norm in defining the best approximation of a function using a family of functions.

1. $J[\alpha h] = \alpha J[h]$ for any function $h \in \mathcal{F}$ and $\alpha \in \mathbb{R}$;

2. $J[h_1 + h_2] = J[h_1] + J[h_2]$ for any $h_1, h_2 \in \mathcal{F}$.

EXAMPLE 4. The integral

$$J[h] = \int_a^b h(x)\,dx$$

defines a linear functional on $\mathcal{D}(a, b)$.

EXAMPLE 5. The integral

$$J[h] = \int_a^b \Big(\alpha_0(x)h(x) + \alpha_1(x)h'(x) + \cdots + \alpha_n(x)h^{(n)}(x)\Big)\,dx,$$

where $\alpha_i(x)$ are functions in $\mathcal{D}(a, b)$, defines a linear functional on $\mathcal{D}_n(a, b)$.

Let $J[y]$ be a functional defined on some normed linear space, and

$$\Delta J[h] = J[y + h] - J[y]$$

be its change due to an increment $h(x)$ of the "variable" $y = y(x)$. When $y(x)$ is fixed, $\Delta J[h]$ is a functional of h. Suppose

$$\Delta J[h] = \phi(h) + \epsilon\|h\|,$$

where $\phi(h)$ is a linear functional and $\|h\| \to 0$ as $\epsilon \to 0$. Then the functional $J[y]$ is said to be *differentiable*. The linear functional $\phi[h]$ which differs from $\Delta J[h]$ by the higher order infinitesimal term $\epsilon\|h\|$, is called the *variation* (or *differential*) of $J[y]$ and denoted by $\delta J[h]$. *This differential is unique when the functional is differentiable.*

Theorem 1 *The differentiable functional $J[y]$ has an extremum at $y = y^*$ only if its variation vanishes for $y = y^*$, that is,*

$$\delta J[h] = 0$$

for all admissible functions h.

In the theorem, *admissible functions* refer to those that satisfy the constraints of a given variational problem. We refer to [1, p. 13] for a proof of Theorem 1.

The function $J[y]$ has an *extremum* for $y = y^*$ if $J[y] - J[y^*]$ does not change its sign in some neighborhood of the curve $y^*(x)$.

3 Euler's Equation

Now we consider what might be called the "simplest" variational problem. Let $F(x, y, z)$ be a function which is twice continuously differentiable with respect to all arguments. Among all functions $y(x)$ that are continuously differentiable over $[a, b]$ and satisfy the boundary conditions

$$y(a) = y_a \qquad \text{and} \qquad y(b) = y_b, \tag{1}$$

find the one for which the functional

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{2}$$

has an extremum.

Suppose we increase $y(x)$ by $h(x)$. For $y(x) + h(x)$ to continue to satisfy the boundary conditions (1), we must have

$$h(a) = h(b) = 0. \tag{3}$$

The corresponding increment in the functional is

$$\begin{aligned} \Delta J &= J[y+h] - J[y] \\ &= \int_a^b F(x, y+h, y'+h')\, dx - \int_a^b F(x, y, y')\, dx \\ &= \int_a^b \Big(F(x, y+h, y'+h') - F(x, y, y') \Big)\, dx \\ &= \int_a^b \Big(F_y(x, y, y')h + F_{y'}(x, y, y')h' \Big)\, dx + \cdots, \end{aligned}$$

after applying Taylor's expansion. The integral in the right hand side of the last equation above represents the principal linear part of the increment ΔJ, hence the variation of $J[y]$ is

$$\delta J = \int_a^b \Big(F_y(x, y, y')h + F_{y'}(x, y, y')h' \Big)\, dx.$$

According to Theorem 1, $J[y]$ has an extremum only if $\delta J = 0$. Thus we have

$$\begin{aligned} 0 &= \int_a^b (F_y(x, y, y')h + F_{y'}(x, y, y')h')\, dx \\ &= \int_a^b h \left(F_y - \frac{d}{dx} F_{y'} \right)\, dx + F_{y'} h|_a^b \\ &= \int_a^b h \left(F_y - \frac{d}{dx} F_{y'} \right)\, dx, \end{aligned}$$

under the conditions (3). Since the function h is arbitrary except $h(a) = h(b) = 0$, it is not difficult to show that the last equation above implies

$$F_y - \frac{d}{dx} F_{y'} = 0. \tag{4}$$

Equation (4) is called *Euler's equation*. Thus, the functional (2) has an extremum for a given function $y(x)$ only if $y(x)$ satisfies Euler's equation. Since the equation is second order, its solution depends on two constants, which can be determined from the two boundary conditions (1). Nevertheless, in some special cases, Euler's equation can be reduced to a first-order differential equation.

Case 1 *The integrand does not depend on* x*.* So the functional has the form

$$\int_a^b F(y, y')\,dx.$$

Euler's equation becomes

$$F_y - \frac{d}{dx} F_{y'}(y, y') = 0.$$

Take the partial derivative:

$$F_y - F_{y'y} y' - F_{y'y'} y'' = 0.$$

Multiplying the above equation by y', we obtain

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

Thus, we end up with a first order differential equation:

$$F - y' F_{y'} = C, \qquad \text{for some constant } C.$$

Case 2 *The integrand does not depend on* y*.* So the functional has the form

$$\int_a^b F(x, y')\,dx.$$

Euler's equation reduces to

$$\frac{d}{dx} F_{y'} = 0,$$

which yields the first-order equation:

$$F_{y'} = C, \qquad \text{for some constant } C.$$

Solving this equation for y', we obtain an equation of the form

$$y' = f(x, C),$$

which is integrated for the curve:

$$y = \int f(x, C)\,dx + D, \qquad \text{for some constant } D.$$

Case 3 *The integrand does not depend on* y'*.* Euler's equation takes the form

$$F_y(x, y) = 0,$$

which is not a differential equation.

Case 4 In a variety of problems, the functional is the integral of a function $f(x, y)$ with respect to the arc length $\sqrt{1+y'^2}dx$. Here $F = f(x, y)\sqrt{1+y'^2}$. Euler's equation can be transformed as follows:

$$\begin{aligned}
0 &= F_y - \frac{d}{dx}F_{y'} \\
&= f_y(x,y)\sqrt{1+y'^2} - \frac{d}{dx}\left(f(x,y)\frac{y'}{\sqrt{1+y'^2}}\right) \\
&= f_y\sqrt{1+y'^2} - f_x\frac{y'}{\sqrt{1+y'^2}} - f_y\frac{y'^2}{\sqrt{1+y'^2}} - f\frac{y''}{(1+y'^2)^{3/2}} \\
&= \frac{1}{\sqrt{1+y'^2}}\left(f_y - f_x y' - f\frac{y''}{1+y'^2}\right),
\end{aligned}$$

that is,

$$f_y - f_x y' - f\frac{y''}{1+y'^2} = 0.$$

Example 6. For the functional

$$J[y] = \int_a^b (x-y)^2\, dx,$$

Euler's equation reduces to the equation

$$x - y = 0.$$

So the solution is the line $y = x$, along which the integral vanishes. This is Case 3.

Example 7. Consider the functional

$$J[y] = \int_1^2 \frac{\sqrt{1+y'^2}}{x}\, dx, \qquad y(1) = 0, \qquad y(2) = 1.$$

The integrand does not contain y. This is Case 2, and Euler's equation has the form $F_{y'} = C$, for some C. Thus,

$$\frac{y'}{x\sqrt{1+y'^2}} = C.$$

so that

$$y'^2(1 - C^2x^2) = C^2x^2.$$

Since C can take on either a positive or a negative value, we obtain the derivative from the above:

$$y' = \frac{Cx}{\sqrt{1-C^2x^2}}.$$

Integrating the above gives us

$$\begin{aligned}
y &= \int \frac{Cx}{\sqrt{1-C^2x^2}}\, dx \\
&= -\frac{1}{C}\sqrt{1-C^2x^2} + D,
\end{aligned}$$

or, equivalently,

$$(y-D)^2 + x^2 = \frac{1}{C^2}.$$

Namely, the solution is a circle centered on the y-axis. From the boundary conditions, we find that

$$C = \frac{1}{\sqrt{5}} \qquad \text{and} \qquad D = 2.$$

Thus, the final solution is

$$(y-2)^2 + x^2 = 5.$$

EXAMPLE 8. Among all the curves joining two points (x_0, y_0) and (x_1, y_1), find the one which generates the surface of minimum area when rotated about the x-axis. In this problem, the functional is the area of the surface of revolution as a result of rotating the curve $y = y(x)$ about the x-axis:

$$2\pi \int_{x_0}^{x_1} y\sqrt{1+y'^2}\, dx.$$

We let $F(y, y') = y\sqrt{1+y'^2}$ and minimize the integral. This is case 1 where Euler's equation yields the first order equation:

$$F - y'F_{y'} = C.$$

Thus, we have

$$y\sqrt{1+y'^2} - y\frac{y'^2}{\sqrt{1+y'^2}} = C,$$

or

$$y = C\sqrt{1+y'^2}.$$

So,

$$y' = \pm\sqrt{\frac{y^2 - C^2}{C^2}},$$

which becomes

$$dx = \frac{C}{\sqrt{y^2 - C^2}} dy,$$

with solution

$$x + D = C \ln \frac{y + \sqrt{y^2 - C^2}}{C},$$

or equivalently,

$$y = C \cosh \frac{x+D}{C}.$$

The solution curve in Example 7 is a *catenary*, and the surface generated by rotation of the catenary is called a *catenoid*. The constants C and D are determined using the conditions

$$y(x_0) = y_0 \qquad \text{and} \qquad y(x_1) = y_1.$$

Figure 2 plots the solution curve $y = 2\cosh(\frac{x+3}{2})$ joining two points $(1, 2\cosh(2))$ and $(2, \cosh(\frac{5}{2}))$.

Depending on the locations (x_0, y_0) and (x_1, y_1), the solution surface may not exist. Consider the case where the distance between the two points is sufficiently large compared to y_0 and y_1. The area generated by rotating the polygonal line $\boldsymbol{p}_1\boldsymbol{p}_2\boldsymbol{p}_3\boldsymbol{p}_4$, where $\boldsymbol{p}_1 = (x_0, y_0)$, $\boldsymbol{p}_2 = (x_0, 0)$, $\boldsymbol{p}_3 = (x_1, 0)$, and $\boldsymbol{p}_4 = (x_1, y_1)$ will be less than the area generated by rotating a smooth curve passing through the points.

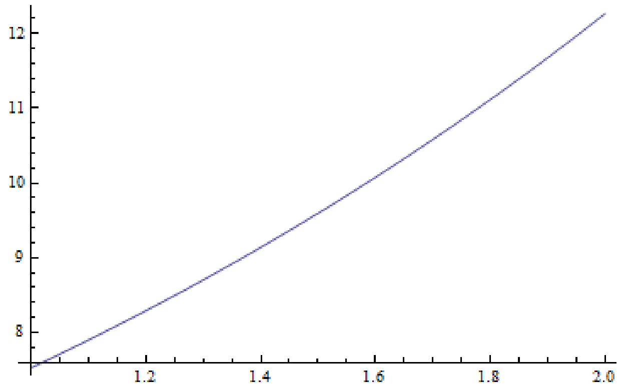

Figure 2: A catenary $y = 2\cosh(\frac{x+3}{2})$ over $[1, 2]$.

References

[1] I. M. Gelfand and S. V. Fomin. *Calculus of Variations* Prentice-Hall, Inc., 1963; Dover Publications, Inc., 2000.

[2] R. Weinstock. *Calculus of Variations: With Applications to Physics and Engineering*. Dover Publications, Inc., 1974.