Calculus of Variations and Further

Yuxuan Chen

June 2022

1 Introduction

In June of 1696, John (also known as Johann or Jean) Bernoulli challenged the greatest mathematicians of the world to solve the following new problem (Bernoulli, 1696; Goldstine, 1980):

> Given points A and B in a vertical plane to find the path AMB down which a movable point M must, by virtue of its weight, proceed from A to B in the shortest possible time[Kot14, Chapter 1].

This is the classic problem of the curve of fastest descent. For its simplest case, let us assume that the particle is moving without friction, and then the law of conservation of mechanic energy holds. Assume that the particle starts its motion from the origin and moves down in the direction of y-axis (downward). We get:

$$\tfrac{1}{2}mv^2 - mgy = 0.$$

Thus, we get

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

To get the total time needed for the particle to descent, we can add up dt. As the velocity is changing constantly, we start by studying a very small time interval dt. We have:

$$dt = \tfrac{dl}{v}.$$

Express dl in x and y, we get:

$$dl = \sqrt{(dx)^2 + (dy)^2} = \sqrt{1 + (\tfrac{dy}{dx})^2}\ dx.$$

Write $\frac{dy}{dx}$ as y', and express v as $\sqrt{2gy}$. Thus we have:

$$dt = \tfrac{\sqrt{1+(y')^2}}{\sqrt{2gy}}\ dx.$$

Finally, we get the total time for descent as T (x_2 stands for the final position):

$$T = \int dt = \frac{1}{\sqrt{2g}} \int_0^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{y}}\, dx$$

Here, we can see y and y' are unknown functions, yet we want to find out the minimum value of T, an integral containing y and y'. This leads to our first definition:

Definition 1. *A functional is an operator that maps functions to real numbers[Kot14, Chapter 2].*

And in the case of the curve of fastest descent, T is a functional that maps functions y and y' to a real number. Note that we have set the starting point as $x = 0$ in order to keep the problem simple, and it is NOT a complete solution to the problem of the curve of fastest descent, nor is the problem my main topic here. This problem is just an introduction to calculus of variations.

In this paper, I'll explore the basic calculus of variations, some further ideas, and the applications of calculus of variations in optics. First, the proof of calculus of variations will be discussed in the context of the simple problem described above. Second, we will go a bit further to Hamilton's principle and other related topics. Then, the background of optics will be discussed. Finally, the solution of problems will be discussed.

Acknowledgement

I would like to thank my math teacher Keke He for his patient explanations and helpful conversations. And I would like to express my gratitude to Simon Rubinstein-Salzedo and my TA Masashi Alex Deweese, without whom I will never make the paper.

2 Proof of Calculus of Variations

In this section, I'm going to explore the standard forms of calculus of variations and solve the simplified problem of the curve of fastest descent. Remember, we have $T = \frac{1}{\sqrt{2g}} \int_0^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{y}}\, dx$. To simplify the problem further (so that we can focus on the maths idea), we can ignore the constant coefficient $\frac{1}{\sqrt{2g}}$, and we can define another functional as:

$$I = \int_0^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{y}}\, dx.$$

Also, to make the writing simpler, here we can write $\frac{\sqrt{1+(y')^2}}{\sqrt{y}}$ as $f(y, y'; x)$ since it is a functional with variations y and y' and is also indirectly relevant to x.

Now, let's start with a more general form of functional. Suppose we have:

$$I = \int_{x_1}^{x_2} f(y, y'; x)\ dx.$$

And suppose we are going to explore the minimum value of I. Note that the maximum can be found in almost the same way.

2.1 Lead into a Parameter

There must exist a specific function y that enable the functional I to reach its minimum. In order to find that specific function y, we can lead into a parameter ϵ. The key point of calculus of variations is that we should take small changes of function y, and we can do that by varying the parameter ϵ.

Suppose the function y for minimum I is $y(0, x)$, where $\epsilon = 0$ and x is the independent variable. To vary y by varying ϵ, we can define y as $y(\epsilon, x) = y(0, x) + \epsilon\eta(x)$, where $\eta(x)$ is an arbitrary continuous function about independent variable x.

Note that we have already assumed that when $\epsilon = 0$ the functional I reaches its minimum, for any other value of ϵ and any non-zero value of function $\eta(x)$, the distortion will lead to a bigger I.

We have to satisfy the boundary conditions, $y(\epsilon, x_1) = y(0, x_1)$ and $y(\epsilon, x_2) = y(0, x_2)$, as the distortion of function y should not affect the initial position A and final position B in this context (recall the description at the beginning of the essay!), and similar boundary conditions should be recognized in other problems of calculus of variations. So, we have boundary conditions:

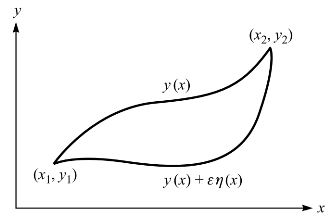
$$\eta(x_1) = 0 \text{ and } \eta(x_2) = 0.$$

2.2 Transform I into a Functional with a Single Parameter

Now, we can write I as $I(\epsilon) = \int_{x_1}^{x_2} f(y(\epsilon, x), y'(\epsilon, x); x)\ dx$. As we change the function y, we change the parameter ϵ only, and I is now transformed to be a functional with a single parameter.

To make the minimum I, as in normal case of functions, we should have extreme when $\epsilon = 0$. Thus, the equation should hold:

$$\frac{\partial I}{\partial \epsilon} \mid_{\epsilon=0} = 0.$$



2.3 Differentiate I

Simplify the equation. As in a classic mechanic system, the partial differentials must be continuous, which makes the order of differential and integral matter little, we have:

$$\frac{\partial I}{\partial \epsilon} = \frac{\partial}{\partial \epsilon}\int_{x_1}^{x_2} f(y(\epsilon, x), y'(\epsilon, x); x)\ dx = \int_{x_1}^{x_2} \frac{\partial}{\partial \epsilon} f(y(\epsilon, x), y'(\epsilon, x); x)\ dx.$$

Then, we can apply chain rule in the partial differential, and we get:

$$\frac{\partial I}{\partial \epsilon} = \int_{x_1}^{x_2} (\frac{\partial f}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial f}{\partial y'}\frac{\partial y'}{\partial \epsilon})\ dx.$$

Recall that $y = y(\epsilon, x) = y(o, x) + \epsilon\eta(x)$, so we have:

$$\frac{\partial y}{\partial \epsilon} = \eta(x).$$

As for $\frac{\partial y'}{\partial \epsilon}$, note that y' stands for $\frac{\partial\ y(\epsilon,x)}{\partial x}$, we can write:

$$\frac{\partial y'}{\partial \epsilon} = \frac{\partial}{\partial \epsilon}(y'(0, x) + \epsilon\eta'(x)) = \eta'(x).$$

So, we get:

$$\frac{\partial I}{\partial \epsilon} = \int_{x_1}^{x_2} (\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y'}\eta'(x))\ dx = \int_{x_1}^{x_2} \frac{\partial f}{\partial y}\eta(x)\ dx + \int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\eta'(x)\ dx.$$

To cancel out $\eta'(x)$, we can expand $\int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\eta'(x)\ dx$ using integral by parts. Thus, let $u = \frac{\partial f}{\partial y'}$ and $dv = \eta'(x)\ dx$, and we can deduce:

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\eta'(x)\ dx = \int_{x_1}^{x_2} u\ dv = [uv]_{x_1}^{x_2} - \int_{x_1}^{x_2} v\ du =$$
$$[\frac{\partial f}{\partial y'}\eta(x)]_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx}(\frac{\partial f}{\partial y'})\eta(x)\ dx.$$

Recall the boundary conditions. We have:

$$[\frac{\partial f}{\partial y'}\eta(x)]_{x_1}^{x_2} = \eta(x_1)\frac{\partial f}{\partial y'}\ |_{x_1} - \eta(x_2)\frac{\partial f}{\partial y'}\ |_{x_2} = 0.$$

Then, we have:

$$\frac{\partial I}{\partial \epsilon} = \int_{x_1}^{x_2} \frac{\partial f}{\partial y}\eta(x)\ dx - \int_{x_1}^{x_2} \frac{d}{dx}(\frac{\partial f}{\partial y'})\eta(x)\ dx = \int_{x_1}^{x_2} [\frac{\partial f}{\partial y} - \frac{d}{dx}(\frac{\partial f}{\partial y'})]\eta(x)\ dx.$$

Since $\eta(x)$ is an arbitrary function, when $\frac{\partial I}{\partial \epsilon} = 0$, we have:

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

And it is the so-called Euler-Lagrange equation. Note that it is only a necessary condition for I to be an extreme value. Thus, the solution of the Euler-Lagrange equation may not yield the minimizing curve. Ordinarily we must verify whether or not this solution yields the curve that actually minimizes the integral, but frequently physical or geometrical considerations enable us to tell whether the curve so obtained makes the integral a minimum or a maximum[CHO03, Chapter 8].

2.4 The Solution to the Curve of Fastest Descent

Recall that:

$$I = \int_0^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{y}}\ dx$$
$$f(y, y'; x) = \frac{\sqrt{1+(y')^2}}{\sqrt{y}}.$$

When the Euler-Lagrange equation holds, we have $\frac{d}{dx}(\frac{\partial f}{\partial y'}) = \frac{\partial f}{\partial y}$.

Since we have

$$\frac{\partial f}{\partial y'} = \frac{\partial}{\partial y'} \frac{\sqrt{1+(y')^2}}{\sqrt{y}} = \frac{1}{\sqrt{y}} \frac{y'}{\sqrt{1+(y')^2}},$$
$$\frac{d}{dx}(\frac{\partial f}{\partial y'}) = \frac{-\frac{1}{2}(y')^2(1+(y')^2)+yy''}{y^{\frac{3}{2}}(1+(y')^2)^{\frac{3}{2}}},$$

and

$$\frac{\partial f}{\partial y} = -\frac{1}{2}\sqrt{1+(y')^2}\ y^{-\frac{3}{2}},$$

we can deduce that:

$$\frac{2y''}{1+(y')^2} = -\frac{1}{y}.$$

In order to solve the above equation, we can make the left hand side and right hand side both differentials of functions. So we can multiply y' to both sides.Let LHS be the left hand side, and RHS be the right hand side. Integral each side, then:

$$\text{LHS} = \int \frac{2y''y'}{1+(y')^2} = \int \frac{1}{1+(y')^2}\ d(1+(y')^2) = \ln(1+(y')^2);$$
$$\text{RHS} = -\int \frac{1}{y}y' = -\ln y + \ln k.$$

So,

$$(y')^2 = \frac{k}{y} - 1.$$
$$y' = \frac{dy}{dx} = \pm\sqrt{\frac{k}{y} - 1}$$

Thus, we can get:

$$\sqrt{\frac{y}{k-y}}\ dy = \pm dx$$

Let $y = \frac{k}{2}(1-\cos u)$, so $dy = \frac{k}{2}\sin u\ du$. We have:

$$\sqrt{\frac{y}{k-y}}\ dy = \sqrt{\frac{\frac{k}{2}(1-\cos u)}{k-\frac{k}{2}(1-\cos u)}}\ \frac{k}{2}\ \sin u\ du = \frac{k}{2}\frac{\sqrt{1-(\cos u)^2}}{1+\cos u}\ \sin u\ du = \frac{k}{2}\frac{(\sin u)^2}{1+\cos u}\ \sin u\ du = \frac{k}{2}(1-\cos u)\ du$$

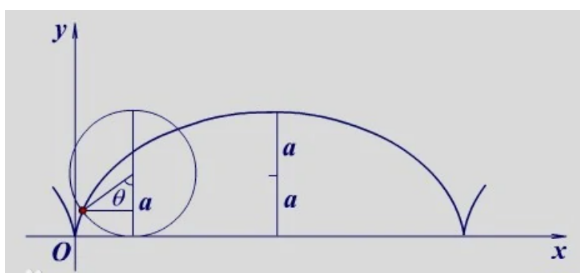
Thus,

$$\frac{k}{2}(1-\cos u)\ du = \pm dx$$
$$\int \frac{k}{2}(1-\cos u)\ du = \pm \int dx$$
$$x = \pm\frac{k}{2}(u-\sin u) + const.$$

Recall the boundary condition that when $y = 0$ i.e., $u = 0$, $x = 0$, so $const = 0$.

In this way, we've found the parameter equations for the curve of fastest descent ($a = \frac{k}{2}$ and $\theta = u$).

$$\begin{cases} x = a(\theta - \sin\theta) \\ y = a(1-\cos\theta) \end{cases}$$



3 Hamilton's principle and canonical equations of motion

For the condition needed for the minimum value of functional I, we should have:

$$\delta I = \delta \int_{t_1}^{t_2} f(y, y', x)\ dx.$$

where δ stands for variations (This approach gives exactly the same result as partial differential).

Definition 2. *The change of value of y caused by the change of form of y while x stays the same is called variation.*

In the context of previous discussions, we can write:

$$\delta y(x_0) = y(\epsilon, x_0) - y(0, x_0).$$

In this way, we can rewrite the boundary conditions as:

$$\delta y(x_1) = 0 \text{ and } \delta y(x_2) = 0.$$

Generalize the functional I with generalized coordinates qs and their time derivatives $\dot{q}$s (generalized velocity). Now we can write I as:

$$I = \int_{t_1}^{t_2} L(q_i, \dot{q}_i; t)\ dt,$$

where i indicates multiple dimensions can be considered (the use of L is explained later in this section).

Here we introduce the generalized momentum ps. We can describe the state of a system by Hamilton's function (or the Hamiltonian) H, by Hamilton's equations or canonical equations:

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$
$$\dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

To make it easier to understand, we can see it in a mechanic context.

Definition 3. *Let T be the kinetic energy in this system, while V be the potential energy in the system, then the Lagrangian $L = T - V$.*

We can deduce from Newtonian mechanics that when $I = \int_{t_1}^{t_2} L\ dt$ is minimized, the motion is valid (since it is loosely related to the topic, I'm not going to give proof, but it is not difficult, you can really try!).

Consider the units. We have L as energy $\times$ time and the generalized coordinates q (as positions in this case). As momentum p is force $\times$ time, and energy is force $\times$ distance, we can get the relation:

$$p_i = \frac{\partial L}{\partial q_i}.$$

Finally, this leads us to define the Hamiltonian as:

$$H = \sum p_i \dot{q}_i - L.$$

As $\dot{q}_i$ can be expressed by q_i, p_i, and t, we can write Hamiltonian as:

$$H = H(q_i, p_i, t).$$

We are ready to deduce Hamilton's equation from Hamilton's principle. The original Hamilton's principle refers to paths in configuration space, so in order to extend the principle to phase space, we must modify it such that the integrand of the action I is a function of both the generalized coordinates and momenta and their derivatives.

Since we have $H = \sum p_i\dot{q}_i - L$, we can rewrite I in this form:

$$I = \int_{t_1}^{t_2} (\sum p_i\dot{q}_i - H(q_i, p_i, t))\ dt.$$

Then we obtain that:

$$\delta I = \delta \int_{t_1}^{t_2} (\sum p_i\dot{q}_i - H(q_i, p_i, t))\ dt = 0.$$

Carrying out the variation (Note the rules are almost the same as differential), we obtain:

$$\int_{t_1}^{t_2} \sum \delta(p_i\dot{q}_i - H(q_i, p_i, t))\ dt = 0$$
$$\int_{t_1}^{t_2} \sum (p_i\delta\dot{q}_i + \dot{q}_i\delta p_i - \frac{\partial H}{\partial q_i}\delta q_i - \frac{\partial H}{\partial p_i}\delta p_i)\ dt = 0.$$

Note that $\dot{q}_i = \frac{d}{dt}q_i$, so $\delta\dot{q}_i = \frac{d}{dt}\delta q_i$.

Now we integrate the term $p_i\delta\dot{q}_i\ dt$ by part.

$$\int_{t_1}^{t_2} \sum p_i\delta\dot{q}_i\ dt = \int_{t_1}^{t_2} \sum p_i\delta\frac{d}{dt}\delta q_i\ dt$$

As $\frac{d}{dt}(p_i\delta q_i) = \dot{p}_i\delta q_i + p_i\frac{d}{dt}\delta q_i$, we have:

$$\int_{t_1}^{t_2} \sum p_i\delta\dot{q}_i\ dt = \int_{t_1}^{t_2} \sum \frac{d}{dt}(p_i\delta q_i)\ dt - \int_{t_1}^{t_2} \sum \dot{p}_i\delta q_i\ dt = [p_i\delta q_i]_{t_1}^{t_2} - \int_{t_1}^{t_2} \sum \dot{p}_i\delta q_i\ dt.$$

As we have boundary conditions as $\delta q_i(t_1) = \delta q_i(t_2) = 0$, we get:

$$\int_{t_1}^{t_2} \sum p_i\delta\dot{q}_i\ dt = -\int_{t_1}^{t_2} \sum \dot{p}_i\delta q_i\ dt.$$

Substituting this back, we have:

$$I = \int_{t_1}^{t_2} \sum p_i\dot{q}_i - H(q_i, p_i, t)\ dt = \int_{t_1}^{t_2} \sum(-\dot{p}_i\delta q_i + \dot{q}_i\delta p_i - \frac{\partial H}{\partial q_i}\delta q_i - \frac{\partial H}{\partial p_i}\delta p_i)\ dt =$$
$$\int_{t_1}^{t_2} \sum[(\dot{q}_i - \frac{\partial H}{\partial p_i})\delta p_i - (\dot{p}_i + \frac{\partial H}{\partial q_i})\delta q_i]\ dt = 0.$$

As we change the form of functions to have a minimum, we can see that both δps and δqs are arbitrary, so the coefficients should be zero.

Thus, we get[CHO03]:

$$\begin{cases} \dot{q}_i - \frac{\partial H}{\partial p_i} = 0 \\ \dot{p}_i + \frac{\partial H}{\partial q_i} = 0 \end{cases}$$

4 Double Integrals

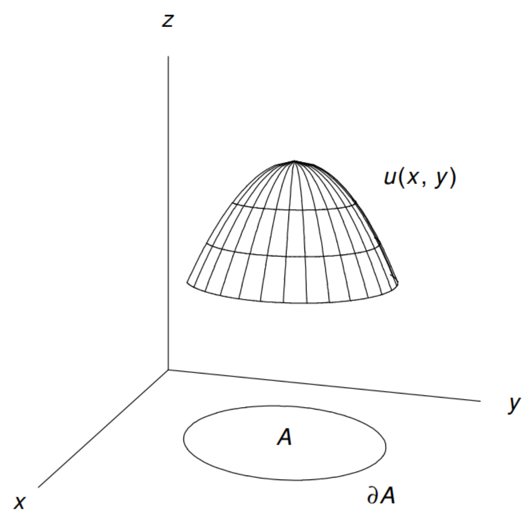
Define a functional as:

$$I(u) = \iint_A F(x, y, u, u_x, u_y)\ dx\ dy.$$

Now we are trying to find the minimum value of functional $I(u)$, where u is a function of both x and y, subject to the boundary condition

$$u(x, y) = u_0(x, y),$$

for all (x, y) on the boundary curve ∂A[Kot14, Chapter 4.8].



For the sake of clarity, we represent $\frac{\partial u}{\partial x}$ as u_x and $\frac{\partial u}{\partial y}$ as u_y. Similar expressions are used for other partial differentials.

Let $\bar{u}$ be the function u with some change in parameters (similar to $y(\epsilon, x)$ in above sections), and we have:

$$\bar{u} = u + \epsilon\eta(x, y),$$

where u stands for the optimal solution that makes $I(u)$ a minimum.

So, we can write $I(u)$ in the following form as:

$$I(u) = \iint_A F(x, y, u + \epsilon\eta, u_x + \epsilon\eta_x, u_y + \epsilon\eta_y)\ dx\ dy = I(\epsilon).$$

Now, let consider the so-called *Total variation*.

$$\Delta I = I(\epsilon) - I(0),$$

which can be also written in form:

$$\Delta I(u) = \iint_A F(x,y,u+\epsilon\eta, u_x+\epsilon\eta_x, u_y+\epsilon\eta_y)\, dx\, dy - \iint_A F(x,y,u,u_x,u_y)\, dx\, dy.$$

Let's now expand the above in a Taylor series in ϵ in the usual way. We obtain:

$$\Delta I = \delta I + \frac{1}{2}\delta^2 I + \dots$$

where

$$\Delta I = \epsilon \iint_A (\frac{\partial F}{\partial u}\eta + \frac{\partial F}{\partial u_x}\eta_x + \frac{\partial F}{\partial u_y}\eta_y)\, dx\, dy.$$

Note that we have:

$$\frac{\partial F}{\partial u_x}\eta_x = \frac{\partial \eta}{\partial x}\frac{\partial F}{\partial u_x}$$

and

$$\frac{\partial}{\partial x}(\eta \frac{\partial F}{\partial u_x}) = \frac{\partial \eta}{\partial x}\frac{\partial F}{\partial u_x} + \eta \frac{\partial}{\partial x}(\frac{\partial F}{\partial u_x}).$$

So, we get:

$$\frac{\partial F}{\partial u_x}\eta_x = \frac{\partial}{\partial x}(\eta \frac{\partial F}{\partial u_x}) - \eta \frac{\partial}{\partial x}(\frac{\partial F}{\partial u_x}),$$

and similarly:

$$\frac{\partial F}{\partial u_y}\eta_y = \frac{\partial}{\partial y}(\eta \frac{\partial F}{\partial u_y}) - \eta \frac{\partial}{\partial y}(\frac{\partial F}{\partial u_y}).$$

So,

$$\delta I = \epsilon \iint_A [\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}(\frac{\partial F}{\partial u_x}) - \frac{\partial}{\partial y}(\frac{\partial F}{\partial u_y})]\eta\, dx\, dy + \epsilon \iint_A [\frac{\partial}{\partial x}(\eta \frac{\partial F}{\partial u_x}) + \frac{\partial}{\partial y}(\eta \frac{\partial F}{\partial u_y})]\, dx\, dy$$

For further discussion, we need to introduce Green's theorem:

Theorem 1. *Let C be a positively oriented, piecewise-smooth, simple closed curve in the plane and let D be the region bounded by C. If P and Q have continuous partial derivatives on an open region that contains D, then:*

$$\int_C P dx + Q dy = \iint_D (\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y})\, dA.$$

In this context, we have:

$$\iint_A (\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y})\, dx\, dy = \int_{\partial A} P dx + Q dy.$$

So we get:

$$\delta I = \epsilon \iint_A [\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}(\frac{\partial F}{\partial u_x}) - \frac{\partial}{\partial y}(\frac{\partial F}{\partial u_y})]\eta\, dx\, dy + \epsilon \int_{\partial A} (-\eta \frac{\partial F}{\partial u_y})\, dx + (\eta \frac{\partial F}{\partial u_x})\, dy.$$

Recall boundary conditions that $\eta(x,y) = 0$ on ∂A and $\eta(x,y)$ is an arbitrary function, we get:

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}(\frac{\partial F}{\partial u_x}) - \frac{\partial}{\partial y}(\frac{\partial F}{\partial u_y}) = 0.$$

This is the Euler-Lagrange equation in double integrals.

5 Higher-order Derivatives

As we see in the above cases, the functional I is only dependent on function y or u and their first derivative. Now let's try to generalize the basic theory to adapt to situations where I is also dependent on y'' [Kot14, Chapter 4.1].

Define functional I as:

$$I = \int_a^b f(y, y', y''; x)\ dx.$$

We can write down the boundary conditions as:

$$y(a) = y_a, y(b) = y_b, y'(a) = y'_a, y'(b) = y'_b.$$

Note that the starting and ending point should be fixed, and as I is dependent on the second derivative, we can find that similar to the classical proof, we should fix y and y'.

Just the same as classical proof, we define $y(\epsilon, x) = y(0, x) + \epsilon\eta(x)$, where $y(0, x)$ gives the extreme value of I. And the boundary conditions here are:

$$\eta(a) = 0, \eta(b) = 0, \eta'(a) = 0, \eta'(b) = 0.$$

We now have:

$$\frac{\partial I}{\partial \epsilon} = \int_a^b (\frac{\partial f}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial f}{\partial y'^{]}}\frac{\partial y'}{\partial \epsilon} + \frac{\partial f}{\partial y''}\frac{\partial y''}{\partial \epsilon})\ dx.$$

Also,

$$\frac{\partial y}{\partial \epsilon} = \eta,$$
$$\frac{\partial y'}{\partial \epsilon} = \eta',$$
$$\frac{\partial y''}{\partial \epsilon} = \eta''.$$

So, we get:

$$\frac{\partial I}{\partial \epsilon} = \int_a^b (\frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial y'}\eta' + \frac{\partial f}{\partial y''}\eta'')\ dx.$$

Just like in the proof of basic case, we now need to cancel out η' and η''. Using integral by parts, we have:

$$\int_a^b \eta' \frac{\partial f}{\partial y'}\ dx = [\eta\frac{\partial f}{\partial y'}]_a^b - \int_a^b \eta \frac{d}{dx}(\frac{\partial f}{\partial y'})\ dx$$

and

$$\int_a^b \eta'' \frac{\partial f}{\partial y''}\ dx = [\eta' \frac{\partial f}{\partial y''}]_a^b - \int_a^b \eta' \frac{d}{dx}(\frac{\partial f}{\partial y''})\ dx$$
$$= [\eta' \frac{\partial f}{\partial y''} - \eta\frac{d}{dx}(\frac{\partial f}{\partial y''})]_a^b + \int_a^b \eta \frac{d^2}{dx^2}(\frac{\partial f}{\partial y''})\ dx.$$

Recall the boundary conditions, we can deduce that:

$$\frac{\partial I}{\partial \epsilon} = \int_a^b [\frac{\partial f}{\partial y} - \frac{d}{dx}(\frac{\partial f}{\partial y'}) + \frac{d^2}{dx^2}(\frac{\partial f}{\partial y''})]\eta \ dx.$$

Because we have $\frac{\partial I}{\partial \epsilon} = 0$ when $\epsilon = 0$ regardless what value of x and $\eta(x)$ (which is arbitrary), we must have:

$$\frac{\partial f}{\partial y} - \frac{d}{dx}(\frac{\partial f}{\partial y'}) + \frac{d^2}{dx^2}(\frac{\partial f}{\partial y''}) = 0.$$

This above proof can be extended to even higher derivatives. For functional

$$I = \int_a^b f(y, y', y'', \ldots, y^{(n)}; x) \ dx,$$

it can be proven that

$$\frac{\partial f}{\partial y} - \frac{d}{dx}(\frac{\partial f}{\partial y'}) + \cdots + (-1)^n \frac{d^n}{dx^n}[\frac{\partial f}{\partial y^{(n)}}] = 0.$$

And this above equation is called Euler–Poisson equation.

6 Minimizing or Maximizing with Constraints

The calculus of variations is also frequently used in questions with extra constraints. It can simplify the problem by reducing the number of equations.

6.1 Method of Lagrange Multipliers

First let's define an operator ∇[CHO03].

Definition 4. *The vector differential operator is an operation that changes a scalar field to a vector field, and it is called a gradient operator.*

$$\nabla = \frac{\partial}{\partial x_1}\hat{e_1} + \frac{\partial}{\partial x_2}\hat{e_2} + \frac{\partial}{\partial x_3}\hat{e_3},$$

where $\hat{e_1}$, $\hat{e_2}$, and $\hat{e_3}$ are the base unit vectors.

The case in two dimensions is very similar, and we'll use the two-dimensional case in the proof here.

Suppose we have a constraint $G(x, y) = 0$ (a curve), and we need to find extreme for $f(x, y) = c$ (a series of straight lines). The extreme will be reached when they are tangent to each other, i.e., the gradients of G and f are parallel. So, we have:

$$\nabla f = \lambda \nabla G.$$

Expand that, we get:

$$f_x\hat{e_1} + f_y\hat{e_2} = \lambda(G_x\hat{e_1} + G_y\hat{e_2}).$$

So, to get the extreme value for f, we just need to solve equations

$$\begin{cases} f_x - \lambda G_x = 0 \\ f_y - \lambda G_y = 0 \\ G(x,y) = 0 \end{cases}$$

We can now define a new function called F

$$F(x, y, \lambda) = f - \lambda G,$$

and the above equations now turn out to be:

$$\begin{cases} F_x = 0 \\ F_y = 0 \\ F_\lambda = 0 \end{cases}$$

So we have $\nabla F = \vec{0}$, i.e., we now convert the question into the search of extreme value of F.

6.2 Solving problems with Constraints

When the method of Lagrange multipliers are used in the calculus of variations, we have:

Theorem 2. *If $x = x(t)$ is a critical function of the functional subject to the constraint, then there exists a constant λ such that x satisfies the Euler-Lagrange equation corresponding to the integrand $L = F - \lambda G$:*

$$\tfrac{d}{dt}((F - \lambda G)_{\dot{x}}) - (F - \lambda G)_x = 0 \text{[Lev14, Chapter 5]}.$$

When applied in different cases, the independent variables may not necessarily be t, and the partial differential may not be expressed as $\dot{x}$.

6.2.1 Generalized Case

Suppose we are trying to calculate the minimum of functional

$$I = \int_{x_1}^{x_2} f(y, y'; x)\ dx$$

with constraint that

$$J = \int_{x_1}^{x_2} g(y, y'; x)\ dx = Const.$$

So, as discussed above, we can define $F = I - \lambda J$, where

$$F = \int_{x_1}^{x_2} [f(y, y'; x) - \lambda g(y, y'; x)]\ dx.$$

To get the extreme value of F, we have the Euler-Lagrange equation:

$$[\tfrac{d}{dx}(\tfrac{\partial f}{\partial y'}) - \tfrac{\partial f}{\partial y}] - \lambda[\tfrac{d}{dx}(\tfrac{\partial g}{\partial y'}) - \tfrac{\partial g}{\partial y}] = 0.$$

6.2.2 The Example

To better illustrate the idea, we now consider a specific example: when the perimeter of a shape is a constant, what shape gives out the maximum area?

First, we need to model the question as functions and functionals. We have our constraint of perimeter:

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

where l is a constant. And we can write the area A as:

$$A = \frac{1}{2}\int_c (xdy - ydx).$$

The proof of the above equation can be done by substituting $x = r\cos\theta$ and $y = r\sin\theta$, where r is the distance from origin and θ is the angle between the line connecting origin and point (x, y) and the x-axis. We know that area $A = \frac{1}{2}r^2\ d\theta$, and we have

$$\frac{dx}{d\theta} = -r\sin\theta = -y$$
$$\frac{dy}{d\theta} = r\cos\theta = x.$$

So, $d\theta = -\frac{dx}{y} = \frac{dy}{x}$. Since $r^2 = x^2 + y^2$, we have:

$$A = \frac{1}{2}\int x^2\ d\theta + y^2\ d\theta = \frac{1}{2}\int x^2\frac{dy}{x} + y^2\frac{dx}{-y} = \frac{1}{2}\int_c (xdy - ydx).$$

Now, back to our question, we have:

$$A = \frac{1}{2}\int_c (xdy - ydx) = \frac{1}{2}\int_c (x\frac{dy}{dx} - y)\ dx = \frac{1}{2}\int_c (xy' - y)\ dx.$$

Applying the method of Lagrange multipliers, we can define:

$$H = A - \lambda s = \int_c [\frac{1}{2}(xy' - y) - \lambda\sqrt{1+(y')^2}]\ dx$$

We must satisfy the Euler-Lagrange equation to get the maximum, so:

$$\frac{d}{dx}(\frac{\partial H}{\partial y'}) - \frac{\partial H}{\partial y} = 0.$$

Substituting the specific functional, we get:

$$\frac{d}{dx}[\frac{1}{2}x - \frac{\lambda y'}{\sqrt{1+(y')^2}}] + \frac{1}{2} = 0.$$

Integral it, we have:

$$x - \frac{2\lambda y'}{\sqrt{1+(y')^2}} = -x + Const.$$

Since λ is a constant, we can let $\lambda_1 = -2\lambda$, so the equation is now:

$$x + \frac{\lambda_1 y'}{\sqrt{1+(y')^2}} = -x + Const$$
$$y' = \pm\frac{x-c_1}{\sqrt{\lambda_1^2-(x-c_1)^2}},$$

where c_1 is another constant.

Again, as $y' = \frac{dy}{dx}$, we integral it:

$$y = \pm\sqrt{\lambda_1^2 - (x-c_1)^2} + c_2$$
$$(x-c_1)^2 + (y-c_2)^2 = \lambda_1^2,$$

where c_2 is another constant.

Here we can see the shape with maximum area is a circle.

7 Background Information of Optics

In the geometric optics, we have the well known theorem of Format Principle.

Theorem 3. *The optical length of the path followed by light between two fixed different points is the global minima. The optical length is the physical length multiplied by the refractive index of the medium.*

Please pay attention to the word global minimum. Remember that the global minimum is the smallest overall value of a set. So, imagine that we have a set, such that its elements are all the possible optical paths from one point to another. These paths have their respective optical length. What the Fermat principle says is that the only physically valid path of our set is the one that has the smallest value of an optical path length.

Mathematically, Fermat's principle can be described as the time T a point of the ray needs to cover a path between the points A and B (Note v stands for the speed of light at this certain media, c stands for the speed of light in vacuum, and s is the distance travelled), given by,

$$T = \int_{t_0}^{t_1} dt = \frac{1}{c}\int_{t_0}^{t_1} \frac{c}{v}\frac{ds}{dt}dt = \frac{1}{c}\int_A^B n\ ds,$$

where n is the index of refraction, $n = \frac{v}{c}$.

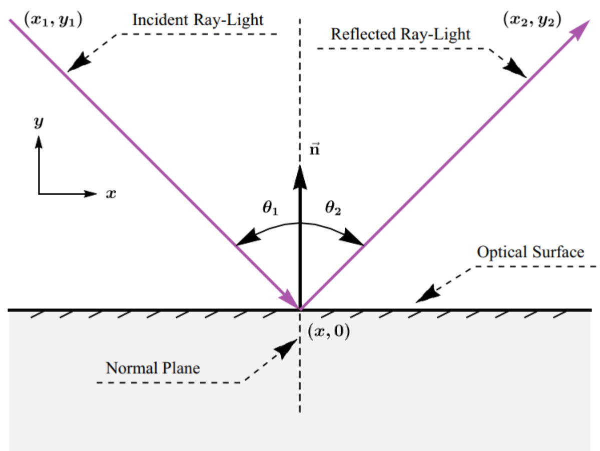
In this way, the minimum of time is converted to the minimum of $\frac{1}{c}\int_A^B n\ ds$, which can be defined as optical path length.

Definition 5. *The optical path length S of a ray from A to B is defined following integral $S = \frac{1}{c}\int_A^B n\ ds$, related to the travel time T by $S = cT$ [RGGA20, Chapter 2.5].*

With the Fermat's principle, we can get strict proof for the laws of reflection and refraction.

7.1 Proof of Law of Reflection

See the diagram below[RGGA20, Chapter 4.3].



We can see if the ray starts from point (x_1, y_1), arrives in a straight line to the point where it is reflected in the surface and then reaches point (x_2, y_2). Let the reflection point be $(x, 0)$.

So, we have the total time travelled as:

$$T = \frac{1}{c}[\sqrt{(x_1 - x)^2 + y_1^2} + \sqrt{(x - x_2)^2 + y_2^2}]$$

To make time T a minimum, we should make $\frac{\partial T}{\partial x} = 0$ and $\frac{\partial^2 T}{\partial x^2} > 0$.

So, we have:

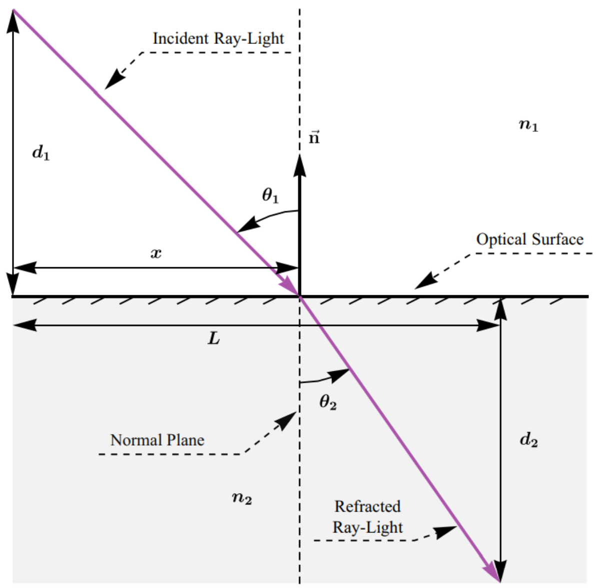
$$\frac{x - x_1}{\sqrt{(x-x_1)^2 + y_1^2}} = \frac{x - x_2}{\sqrt{(x-x_2)^2 + y_2^2}}.$$

Thus, as shown in the graph, $\sin\theta_1 = \sin\theta_2$, i.e., $\theta_1 = \theta_2$.

Similarly, we can prove that $\frac{\partial^2 T}{\partial x^2} > 0$. Since it is not the main topic of this paper, this part of proof is omitted.

7.2 Proof of Law of Refraction

See the diagram below[RGGA20, Chapter 4.4].



From the figure we can put the trajectory of light as,

$$R(x) = n_1\sqrt{d_1^2 + x^2} + n_2\sqrt{d_2^2 + (L - x)^2},$$

where d_1 is the height of the initial position of the light ray. x is the horizontal distance from the initial position of the light ray to the origin of the coordinate system. L is the horizontal distance from the initial position to the final position. d_2 is the height of the final position of the light ray.

So, we can get:

$$\frac{d\ R(x)}{dx} = n_1\frac{x}{\sqrt{x^2+d_1^2}} + n_2\frac{L-x}{\sqrt{d_2^2+(L-x)^2}} = n_1 \sin\theta_1 - n_2 \sin\theta_2.$$

According to Fermat's Principle, we should have $n_1 \sin\theta_1 = n_2 \sin\theta_2$.

8 Applications in Optics

Let the movement of the light ray $q = q(\tau)$ in the $(\tau, q) - plane$. Let $n(q)$ be the index of refraction at the point q[EZ04].

We shall have the length of the light path as $\int_{\tau_0}^{\tau_1} n(q)\ \sqrt{1 + (q')^2}\ d\tau$, and to satisfy Fermat's principle, the functional $\frac{1}{c}\ \int_{\tau_0}^{\tau_1} n(q)\ \sqrt{1 + (q')^2}\ d\tau$ should be a minmum.

8.1 The Euler-Lagrange Equation

If we introduce $L(q, q', \tau) = \frac{n(q)}{c}\sqrt{1 + (q')^2}$, we can apply the conclusion of calculus of variations to get the condition needed to satisfy Fermat's principle:

$$\frac{d}{d\tau}(\frac{\partial L}{\partial q'}) - \frac{\partial L}{\partial q} = 0.$$

In this case, we have:

$$\frac{\partial L}{\partial q'} = \frac{n(q)}{c} \times \frac{1}{2}\frac{2q'}{\sqrt{1+(q')^2}} = \frac{n(q)}{c}\frac{q'}{\sqrt{1+(q')^2}}$$

and

$$\frac{\partial L}{\partial q} = \frac{\sqrt{1+(q')^2}}{c}\frac{\partial n(q)}{\partial q}.$$

Thus we have the Euler-Lagrange equation in the optical case as:

$$\frac{d}{d\tau}\frac{nq'}{\sqrt{1+(q')^2}} = \sqrt{1 + (q')^2}\frac{\partial n}{\partial q}.$$

To simplify notations we choose the units of measurement so that $c = 1$.

8.2 The Hamiltonian Canonical Equations in Optics

As discussed in the previous part, we can extend the generalized momentum in optics. Note that it is not the same unit as in mechanics, but the math is still valid. So we have:

$$p = \frac{\partial L}{\partial q'} = \frac{n(q)}{c}\frac{q'}{\sqrt{1+(q')^2}}.$$

Since $H = pq' - L$, we can have:

$$H = \frac{n(q)}{c}\frac{q'}{\sqrt{1+(q')^2}}q' - \frac{n(q)}{c}\sqrt{1 + (q')^2} = -\frac{n(q)}{c}\frac{1}{\sqrt{1+(q')^2}}.$$

As we have adapted the units so that $c = 1$, we have $H = -\frac{n(q)}{\sqrt{1+(q')^2}}$.

So:

$$H = -\sqrt{\frac{(n(q))^2}{1+(q')^2}}$$
$$H = -\sqrt{(n(q))^2(\frac{1+(q')^2-(q')^2}{1+(q')^2})}$$
$$H = -\sqrt{(n(q))^2 - p^2}.$$

As discussed, the Hamiltonian canonical equations are $q' = \frac{\partial H}{\partial p}$ and $p' = -\frac{\partial H}{\partial q}$. So that we have:

$$q' = -(-\frac{1}{2}\frac{2p}{\sqrt{(n(q))^2-p^2}}) = \frac{p}{\sqrt{(n(q))^2-p^2}}.$$
$$p' = -(-\frac{1}{2}\frac{2n(q)\frac{\partial n(q)}{\partial q}}{\sqrt{(n(q))^2-p^2}}) = \frac{\frac{\partial n(q)}{\partial q}n(q)}{\sqrt{(n(q))^2-p^2}}$$

This a system of ordinary differential equations of the first order.

8.3 Example: Fiber Optics

An optical fiber is a cylindrical waveguide made of a low-loss material such as silica glass. Light is guided through a central core that is embedded in an outer cladding. This cladding has lower refractive index than the core. Light rays that graze the core-cladding boundary at a shallow angle undergo total internal reflection and are guided through the core.

Conventional fibers have constant refractive indices in the core and the cladding and are known as step-index fibers. There are also graded-index fibers that have a refractive index that decreases continuously from its center of the fiber. What is the path of light in a graded-index fiber?[Kot14]

Let's use the cylindrical coordinates to deal with the problem. We have z as the coordinate along the axis of the fiber. According to Fermat's principle, the path connecting two arbitrary points, (r_1, θ_1, z_1) and (r_2, θ_2, z_2), makes the optical path length:

$$l = Tc = \int n(r)\sqrt{dr^2 + (r\ d\theta)^2 + dz^2}.$$

This light path length should be a stationary point. We can choose z as the independent variable, and we get:

$$l = Tc = \int_{z_1}^{z_2} n(r)\sqrt{(r')^2 + (r\theta')^2 + 1}\ dz,$$

where

$$\begin{aligned} r' &= \tfrac{dr}{dz} \\ \theta' &= \tfrac{d\theta}{dz} \end{aligned}$$

The problem is now converted to seeking the the stationary point of functional l. Let $f(r, r', \theta, \theta'; z) = n(r)\sqrt{(r')^2 + (r\theta')^2 + 1}$. Recall the Euler-Lagrange equation, we have:

$$\begin{cases} \frac{\partial f}{\partial r} - \frac{d}{dz}(\frac{\partial f}{\partial r'}) = 0 \\ \frac{\partial f}{\partial \theta} - \frac{d}{dz}(\frac{\partial f}{\partial \theta'}) = 0 \end{cases}$$

So there we have:

$$\frac{\partial f}{\partial \theta'} = n(r) \times \frac{1}{2} \times \frac{2r^2\theta'}{\sqrt{(r')^2+(r\theta')^2+1}} = n(r)\frac{r^2\theta'}{\sqrt{(r')^2+(r\theta')^2+1}}$$

In addition, since z does not appear explicitly in our integrand, we have:

$$f - r'\frac{\partial f}{\partial r'} - \theta'\frac{\partial f}{\partial \theta'} = \frac{n(r)}{\sqrt{(r')^2+(r\theta')^2+1}}$$

These two coupled, first-order, ordinary differential equations enable us to determine $r(z)$ and $\theta(z)$ and to determine the path of light through a graded-index optic fiber.

if $\theta'(0) = 0$, then $\frac{\partial f}{\partial \theta'} = 0$ and $\theta'(z) = 0$ for all z. Rays then remain within a constant θ plane that passes through the axis of symmetry; these rays are known as *meridional* rays. For $\theta' = 0$, equations then simplify significantly.

References

[CHO03] TAI L. CHOW. *Mathematical Methods for Physicists: A concise introduction.* Cambridge University Press, 2003.

[EZ04] H.R. Schwarz E. Zeidler, W. Hackbusch. *Oxford Users' Guide to Mathematics.* OXFORD UNIVERSITY PRESS, 2004.

[Kot14] Mark Kot. *A First Course in the Calculus of Variations.* American Mathematical Society, 2014.

[Lev14] Mark Levi. *Classical Mechanics With Calculus of Variations and Optimal Control: An Intuitive Introduction.* American Mathematical Society, Mathematics Advanced Study Semesters, 2014.

[RGGA20] He´ctor A Chaparro-Romo Rafael G Gonza´lez-Acun˜a. *Stigmatic Optics.* IOP Publishing, Bristol, UK, 2020.