Functionals, the Calculus of Variations, and the Principle of Least Action

Yotam Ofek

May 18, 2017

1. **Introduction**

The calculus of variations is a field of mathematical analysis focused on the optimization of functionals. Though its origin can be traced back as far as the late 17th century, to Bernoulli’s brachistochrone problem, it was not until mathematicians such as Leonhard Euler and Joseph-Louis Lagrange developed more substantive methods and systems of understanding that the calculus of variations became an established field of mathematical study.

The calculus of variations serves as a valuable tool with a wide range of applications such as in the study of topology or in solving boundary value problems (like the Dirichlet Principle or Laplace’s Equation). Its usefulness also extends to physical problems in the form of Lagrangian Mechanics, optics, and determining geodesics.

In this paper I will attempt to provide an introduction to some of the core tenets of the calculus of variations, as well as provide a simple example of its use and a path leading to its application in physical problems.

1. **Optimizing Functionals**

Functionals are, put simply, ‘functions of functions’.

In the same way that a function is a mapping from a variable to a value, a functional is a mapping of function to a value.

In many ways functionals are far more complex than functions (in part because the domain of a functional is far more difficult to work with, let alone visualize). However, this should not scare one away from these incredible devices, as we are already familiar with a number of them.

Perhaps the most common examples of functionals are the definite integral and the inner product.

Both of these produce a number when given a function as an argument.

Suppose then that, out of our desire to analyze a functional, we wished to find its extrema. By now we are well versed with how this is done for a *function*: a function is said to have a maximum at if for all in a neighborhood of we had that for all (or, alternatively, a minimum if ). As explained in Hancock’s book [1], we can apply a similar criteria to determine the maxima and minima of functionals by establishing the following

|  |
| --- |
| Image result for calculus of variations |
| Figure 1. |

Suppose we have a functional . We say that F attains its maximum when its argument is the curve if for all curves, , in a neighborhood of we have that . But what does it mean for a function to be in the neighborhood of another function? If we define another function that satisfies for all x, then define *f* as , we get that . is thus uniformly close to for all values of x. We take this to be the definition of *f* being in a neighborhood of *u* (see fig. 1).

Interpreting this, we see that F having a maximum at *u* means that that any small perturbation to the curve would decrease F.

1. **Example Problem**

A common application of the calculus of variations is in determining the least distance path between two points. An excellent demonstration of this can be found in Bliss’ paper [2].

Suppose we wish to find the shortest path between two points, (x1, y1) and (x2, y2), in . Let y(x) be a continuous, piecewise-smooth function that defines a curve connecting these two points such that.

We know already that the length of a curve, s, is given by the functional

Suppose that y(x) IS the minimum length curve. If we then consider some constant , and some continuous, piecewise-smooth function n(x) that satisfies . Then represents a family of functions (parameterized by ) which satisfy the boundary conditions at x1 and x2 (including y(x) for ), and that are in a neighborhood of y(x) for small .

We thus have that

Noting that y(x) is the path that minimizes s, s must reach a minimum with respect to at 0.

Which we get when we take the derivative under the integral. Note that must be piecewise continuous because we stipulated that y is piecewise smooth.

**LEMMA:** Let M(x) be continuous on the interval [x1, x2]. If the integral

Vanishes for any choice of piecewise-smooth n(x) on [x1, x2] for which n(x1) = n(x2) = 0, then M(x) is a constant.

**PROOF:** If the integral vanishes then that implies



For any constant C (due to the boundary conditions of n). A particular function for n(x) defined by



is clearly zero at x=x1. For it to be zero at x=x2, C and M must satisfy



Now we have that the n(x) from (2) (with the added constraint of (3)) must also satisfy (1). Taking n’(x), we get that everywhere except a finite number of points where M is discontinuous. Putting this into (1) we get



from which the lemma is an immediate consequence.

QED

By our lemma, we have that , which we can solve for y’ to get that y’ is equal to some constant. This means that y(x) must be a straight line

The shortest distance path between two points is the straight line path between them

1. **Lagrangian Mechanics**

In physics there exists an incredibly important functional; one from which all of the principles of classical mechanics can be derived. This functional is called “Action”, and it is given by

where S is the action associated with a given motion defined by a function L (known as the Lagrangian). The Lagrangian is determined by the specific system we choose to analyze, but note that it itself is a function of the position, y(t), and it’s rate of change, y’(t) (L can also be expanded to include as many coordinates as are necessary to completely describe the motion). The motion is considered over a time interval from t1 to t2, so again we have the boundary conditions that y is well defined at the end points (let and ).

The core of this functional’s power stems from one of the most important laws in physics: the principle of least action. This principle states that the motion that a system will execute going from t1 to t2 will be the one which minimizes action. Knowing this, we can employ the calculus of variations to determine what restrictions this imposes on the system.

As demonstrated by Landau [3], to do so we again assume we have found a trajectory, y(t), that minimizes action. Again consider a function n(t) just like in Section III (which vanishes at the boundaries) and consider some constant that is arbitrarily small. Define the function , which we will call the ‘variation of y’ because is a trajectory that differs from y by only an arbitrarily small amount.

The difference in S when we go from to is given by

This difference can be expanded in powers of and inside the integral. This expansion has leading terms of the first order. The requirement for S to have a minimum is that these terms should be zero, so that . Resulting in that

Because , we can integrate by parts to get that

Due to the boundary conditions of , the quantity in the square bracket becomes zero, meaning the integral must also evaluate to zero. However, this must be the case regardless of choice of , which is only possible if

This is known as the Euler-Lagrange equation, and with it we can derive any of the laws of classical mechanics

**EXAMPLE:** Derive Newton’s laws of motion

It is known from other computations that the Lagrangian of a physical system is given by



where K is the kinetic energy in the system and U is the potential energy in the system.

Consider a simple particle of mass m traveling along the x-axis within a potential field U(x).

The kinetic energy of the particle is given by . From this we can write down the Lagrangian



By applying the Euler-Lagrange equation to the Lagrangian we get



Those with a minor degree of physics background will quickly identify the quantity as being the force applied to the particle when at position x. Thus we get that the force applied on the particle is equal to its mass times the second derivative of its position (it’s acceleration).

. We have found Newton’s Laws of motion.

QED

1. **The** **Brachistochrone Problem**

The brachistochrone problem was first proposed by Johann Bernoulli in 1696, and it was in the search for its solution that the calculus of variations was born. Its statement is as follows:

*A small ball moving in a 2-dimensional plane starts at rest and is made to roll from a high point , down to a lower point . What path must it take to reach the end in the shortest possible time?*

Johnson’s paper [4] offers a solution of this problem using Taylor series. However, for the sake of keeping things interesting, and in order to apply the knowledge I have already demonstrated, I will be presenting an alternate proof that instead relies on the Euler Lagrange equation.

|  |
| --- |
|  |
| Figure 2. |

First of all, we must define our coordinate system. For convenience, we will take the downwards direction to represent the positive y direction and the direction going from 0 to to be the positive x direction (so that and ). Now we can represent the path of the ball by a function y(x) on the interval , which must satisfy the boundary conditions and (see fig. 2).

In a standard gravitational field of magnitude g, the potential energy of the ball would be given by , where the zero potential energy point is arbitrarily made to coincide with y=0. Since the ball starts from rest, its initial kinetic energy, K, is zero, and thus its total energy, E, is zero starting at (0,0). Then, by the conservation of energy, we have that at all points during the motion, yielding that . Noting that , we find that the speed, v, is given by

An element of the distance traversed, , would be given by

Now we use the fact that the speed is the rate of distance traveled over time () to get that

Integrating these time elements over the course of the whole motion, get the total time

Note that we have now found T as a functional of y, meaning that paths defined by different y(x) functions will yield different travel times in accordance with this equation. Also note that we can rewrite this as a functional of F, where F is a function of a coordinate (y), its derivative with respect to a parameter (y’), and the parameter (x). This format should seem very familiar seeing as we only encountered it one section ago. Indeed, this functional is of identical form to the action functional, meaning that to minimize time, F must satisfy the same requirements we found that the Lagrangian must satisfy to minimize action.

This means that F satisfies the Euler Lagrange equation.

Proceeding is a matter of computation, though we can simplify it by observing that F has no explicit dependence on x. From the Euler-Lagrange equation we see that the lack of explicit x dependence means that

|  |
| --- |
| Taking (a,b)=() |
| Figure 3. |

where C is some constant. Using our given function for F, we get

which when put in the equation above yields

We can then simplify this equation until we have

where A is a new constant. This is a nonlinear differential equation whose solution is most easily expressed in parametric form.

The solution is

|  |
| --- |
|  |
| Figure 4. |

To satisfy the boundary conditions, t0 is given by

These equations for x(t) and y(t), which define the shortest time path that the ball can travel (see fig. 3), also have an additional geometric significance, as they are the equations of a cycloid. A cycloid is the path traced out by a point on the rim of a wheel as it rolls (see fig. 4), and it is fascinating to think that this particular geometry produces the fastest possible path to roll downhill.

1. **Conclusion**

The calculus of variations is a fascinating branch of analysis, and what I have demonstrated here hardly scratches the surface of its potential.

For instance, I used it to demonstrate that the shortest length path between two points in free space is a line. This result is unsurprising, but the technique I employed lays the groundwork for solving much more interesting problems. If, instead of free space, I had specified that the path is restricted to being along the surface of some shape, then the right answer might no longer be a line, but the same method still works to find it. This is the stem from which the study of geodesics branches off.

Also, as a physics enthusiast, the application of the action functional are seemingly limitless, and it would be a great personal challenge to delve deeper into its uses and mathematical underpinnings.

Finally, the fascinating geometry of the brachistochrone leads to more interesting properties worth investigating. For instance, the brachistochrone is also known as the ‘tautochrone’, meaning it is the curve on which a ball will roll to the bottom in the same amount of time regardless of where on it starts out. Figuring out why this works is another avenue for further investigation.

**References**

[1] Hancock, Harris. “The Calculus of Variations.” *Annals of Mathematics*, vol. 9, no. 1/6, 1894, pp. 179–190. *JSTOR*, [www.jstor.org/stable/1967518](http://www.jstor.org/stable/1967518).

[2] Bliss, Gilbert Ames. “SHORTEST DISTANCES.” *Calculus of Variations*, 1st ed., vol. 1, Mathematical Association of America, 1925, pp. 17–40. *JSTOR*, [www.jstor.org/stable/10.4169/j.ctt5hh931.5](http://www.jstor.org/stable/10.4169/j.ctt5hh931.5).

[3] Landau, Lev Davidovich “Mechanics.” Mechanics, 3rd ed, vol 1, Institute of Physical Problems, USSR Academy of Sciences, Moscow, 1960, pp. 2-4.

[4] Nils P. Johnson. “The Brachistochrone Problem.” *The College Mathematics Journal*, vol. 35, no. 3, 2004, pp. 192–197. *JSTOR*, www.jstor.org/stable/4146894.