CALCULUS OF VARIATIONS

ANDREW CHANG

Abstract. We examine the calculus of variations, first reviewing function extrema and then explaining key results related to finding functional extrema, such as the Fundamental Lemma and the Euler–Lagrange Equation. The connections to physics, particularly the action potential and Lagrangian mechanics, are another focus. We discuss key applications including the brachistochrone problem, minimal surfaces, and geodesics, concluding with a summary of more complex extensions like different boundary conditions and the second variation.

1. Introduction

The calculus of variations is a field in analysis with many applications, centered around using tiny changes in functions, called variations, to find extrema of functionals — "functions of functions" that map a set of functions to real numbers. The functionals we will look at are in the form of a definite integral related to a function and its derivative, e.g. giving the arclength of a curve. By formalizing the techniques used to solve for critical points in these scenarios, a wide range of optimization problems become simple.

A simple but illustrative example is finding the shortest curve between two points. For example, say we have two points $\mathbf{a} = (a, \alpha)$ and $\mathbf{b} = (b, \beta)$ in $\mathbb{R}^2$, and we want to find the path with minimal length. Obviously, this is a straight line, but proving this is nuanced. The arclength, which is an integral that depends on the function of the curve and its derivative, has to be minimized over *all* possible curves between the two points, which is not an easy task at first glance. To solve this problem, we will develop analytical tools that can find the critical points of the arclength, and thus the minima, along all possible curves.

In the following sections we will go over the most important results and applications in the calculus of variations, based roughly on notes by Olver ([3]) and Figueroa-O'Farrill ([2]) in addition to the book by Clegg ([1]).

2. The Euler—Lagrange Equation

We now define the setup that will let us generalize this kind of situation to the fundamental problem we examine. Here, the functional to minimize depends on only the function and its first derivative. The basic condition on our function is that it belongs to C^1 over a relevant interval, meaning it is continuous and has a continuous first derivative.

Our goal is to find the function $y = u(x) \in \mathrm{C}^1[a, b]$ that minimizes the *objective functional*

$$J[u] = \int_a^b L(x, u, u')\, dx,$$

where $L(x, u, u')$ is a given function known as the Lagrangian. We assume that this function is sufficiently smooth so that we can take the derivative of it as needed. We also have boundary conditions at the endpoints of the interval: for now we fix $u(a) = \alpha$ and $u(b) = \beta$. In a later section we will look at different conditions that may occur.

Recall that in multivariable calculus, a critical point of a function f occurs when for a point $x \in \mathbb{R}^n$, $\nabla f(x) = 0$. This is equivalent to saying $\nabla f(x) \cdot v = 0$ for a direction vector v, i.e. the directional

Date: June 6, 2022.

derivative of the function along any direction is equal. This is often written as

$$\frac{d}{d\varepsilon} f[x+\varepsilon v]\bigg|_{\varepsilon=0} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f(x+\varepsilon v) - f(x)) = 0.$$

Analogously, for a functional $J[u]$ we set

$$\frac{d}{d\varepsilon} J[u+\varepsilon v]\bigg|_{\varepsilon=0} = 0.$$

The function v gives the direction of the derivative and is called the *variation* of the function u (sometimes written as δu): similar to a directional derivative vector from point x to a nearby point, we may understand it as a difference of two "neighboring" curves $v(x) = u(x) - u_0(x)$. Thus, $v(x)$ is a differentiable function as well from a to b, with specifically $v(a) = v(b) = 0$ since these are fixed endpoints.

Now substituting and differentiating under the integral, we have

$$\begin{aligned} 0 &= \int_a^b \frac{d}{d\varepsilon} L(x, u+\varepsilon v, u' + \varepsilon v')\bigg|_{\varepsilon=0} dx \\ &= \int_a^b \left(\frac{\partial L}{\partial u} v + \frac{\partial L}{\partial \dot{u}} v' \right) dx \end{aligned}$$

by applying the multivariable chain rule to differentiate. The integral on the right hand side is known as the *first variation* of the functional. We can further simplify this, using integration by parts and noting that $v(a) = v(b) = 0$:

$$\begin{aligned} 0 &= \int_a^b \left(\frac{\partial L}{\partial u}\, v \right) dx + \int_a^b \left(\frac{\partial L}{\partial \dot{u}}\, \dot{v} \right) dx \\ &= \int_a^b \left(\frac{\partial L}{\partial u}\, v \right) dx + \left[\frac{\partial L}{\partial \dot{u}}\, v \right]_{x=a}^{x=b} - \int_a^b \left(\frac{d}{dx} \frac{\partial L}{\partial \dot{u}}\, v \right) dx \\ &= \int_a^b \left(\frac{\partial L}{\partial u} - \frac{d}{dx} \frac{\partial L}{\partial \dot{u}} \right) v\, dx. \end{aligned}$$

By the fundamental lemma, this implies the differential equation

$$\frac{\partial L}{\partial u} = \frac{d}{dx} \frac{\partial L}{\partial \dot{u}}.$$

This is the *Euler–Lagrange equation*, developed by the two in collaboration as a method to find the tautochrone, a curve where a bead lands at the end in the same time no matter where it is placed. This problem is closely related to the brachistochrone problem we will look at later.

3. The Fundamental Lemma

This result is simple, but it will turn out to be useful in many of our proofs (including the line on a plane example from earlier).

Lemma 3.1. *If $f(x)$ is continuous on the closed interval $[a, b]$, and*

$$\int_a^b f(x)\, v(x)\, dx = 0$$

for every continuous function $v(x)$, then $f(x) \equiv 0$ (i.e. it is identically zero) for all $a \le x \le b$.

Proof. A trivial proof can be done by setting $v(x) = f(x)$ so the integral equals $\int_a^b f(x)^2\, dx$, from which the result follows by the trivial inequality. However, a more general proof is possible, as shown below:

Suppose that $f(x) \neq 0$ for some $a < x_0 < b$. Without loss of generality, suppose $f(x) > 0$ there (since we can reverse the sign of $v(x)$ as is appropriate). Then by continuity we know that $f(x) > 0$ for an interval $(x_0 - \varepsilon,\, x_0 + \varepsilon)$ with $\varepsilon > 0$. We may now choose $v(x)$ to be a continuous function that is greater than 0 within this interval and equal to 0 outside it (a "bump function"). An easy example is the piecewise linear function

$$v(x) = \begin{cases} \varepsilon - |x - x_0| & \text{if } |x - x_0| < \varepsilon, \\ 0 & \text{otherwise,} \end{cases}$$

and a smooth example is

$$v(x) = \begin{cases} \varepsilon \exp\left[1 - \left(\varepsilon^{-2}(x - x_0)^2 - 1\right)^{-2}\right] & \text{if } |x - x_0| < \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Figure 1. The linear and smooth examples of bump functions.

The latter allows us to extend the fundamental lemma to when the equality is given for infinitely differentiable $v(x)$. Proceeding from here, we have $f(x)\,v(x) \geq 0$ everywhere, so $\int_a^b f(x)\,v(x)\,dx \geq 0$ and equals 0 if and only if $f(x)\,v(x) \equiv 0$ — a contradiction. ■

4. Hamilton's Principle of Least Action

As it turns out, much of physics can be expressed in terms of the action and Lagrangian that we have just considered. The *principle of least action*, which was formulated by various scientists including Fermat, Euler, and Maupertuis over the years, states that the action, which measures the change in a system, tends to be minimized. (To be precise, the action may often be at any stationary point in its range, not just a minima, so the term *principle of stationary action* is also favored.) This basic law can predict the trajectories that objects naturally take: Maupertuis summarized it with the quote "Nature is thrifty in all its actions". Applying the notion of functionals to formalize this, we obtain *Hamilton's principle*, named for William Rowan Hamilton who reformulated the principle of least action in this way.

Theorem 4.1 (Hamilton's principle)**.** *The true movement of a system from one configuration to another between times t_0 and t_1 is such that the action functional given by the integral*

$$\int_{t_0}^{t_1} L(x(t),\, \dot{x}(t))\,dt$$

is at a stationary point, where the Lagrangian is defined as the difference between the kinetic and potential energies $L(x, \dot{x}) = T(\dot{x}) - V(x)$. The kinetic energy is defined as $T(\dot{x}) = \frac{1}{2}m\|\dot{x}\|^2$, while the potential energy may vary depending solely on the position x.

The integral above is often written succinctly as $\int_{t_0}^{t_1} (T - V)\,dt$. Applying the Euler–Lagrange equation here, since

$$\frac{\partial L}{\partial x} = -\frac{\partial V}{\partial x} \quad \text{and} \quad \frac{\partial L}{\partial \dot{x}} = m\dot{x}$$

the Euler–Lagrange equation reduces to

$$m\ddot{x} = -\frac{\partial V}{\partial x}$$

which is equivalent to Newton's second law, the right hand side being the force related to the potential. For example, for a particle in free fall we have $T = \frac{1}{2}m\dot{y}^2$ and $V = mgy$, which gives $m\ddot{y} = mg$ or $a = g$. Taking the antiderivative twice gives the general solution $y = \alpha + \beta t + \frac{1}{2}gt^2$ with constants α, β.

Describing motion in this way is the basis of Lagrangian and Hamiltonian mechanics, abstractions of classical mechanics beyond Newton's laws that are key to modern physics.

5. Geodesics

We now return to the problem considered originally, and the problem of *geodesics* in general — curves that generally represent the shortest path between two points. First we look at the case of curves in the plane. Here we may define the Lagrangian as $L(x, u, u') = \sqrt{1 + u'^2}$ so that the functional to minimize is the arclength of the curve $u(x)$, given by

$$J[u] = \int_a^b \sqrt{1 + u'^2}\, dx.$$

The Euler–Lagrangian equation gives then

$$0 = -\frac{d}{dx} \frac{u'}{\sqrt{1 + u'^2}} = -\frac{u''}{(1 + u'^2)^{3/2}},$$

and since the denominator is nonzero this gives the simple differential equation $u'' = 0$, of which the solutions are straight line functions $u = cx + d$. Of course, to satisfy the boundary conditions we must choose our constants. The only critical function, and thus potential minimizer, is the line

$$y = \frac{\beta - \alpha}{b - a}(x - a) + \alpha$$

as we would probably expect.

There are more surfaces that can be considered with the same techniques. For example, for any surface given by the graph of a function $z = F(x, y)$, we may find a geodesic curve joining $(a, \alpha, F(a, \alpha))$ and $(b, \beta, F(b, \beta))$ that is parametrized by $y = u(x), z = F(x, u(x))$, by minimizing the functional

$$\begin{aligned} J[u] &= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}\, dx \\ &= \int_a^b \sqrt{1 + \left(\frac{du}{dx}\right)^2 + \left(\frac{\partial F}{\partial x}(x, u(x)) + \frac{\partial F}{\partial u}\frac{du}{dx}\right)^2}\, dx, \end{aligned}$$

subject to boundary conditions $u(a) = \alpha, u(b) = \beta$. There are also alternative configurations for cylindrical and spherical coordinates, where similar techniques may apply. However, such cases require generally convoluted computations and differential geometry techniques may be more helpful in general. We will now proceed to another popular and illustrative problem.

6. The Brachistochrone Problem

Here is another example that was one of the early motivations for developing the calculus of variations. Consider a bead that rolls down a curve between two fixed points, moved by gravity. There are many different curves that work, and the bead will roll down them at different rates. A straight line from the starting point to the end point is simple, but some thought shows that going nearly straight down, then straight across is faster. Several possibilities are shown in Figure 2. There are infinite curves to consider, so how does one find the curve that gives the shortest time? (The word brachistochrone is derived from the Ancient Greek words brákhistos = shortest and khrónos = time.)

To solve this problem, we define the time it takes for the bead to move from start to finish as a functional. First assume the motion occurs in the xy plane, with horizontal displacement x and vertical displacement y. The shape of the wire is then given by a function $y = u(x)$, where $u(0) = (0)$ and $u(l) = h$ are our end points. Let s be the length along the wire from the origin to a point (x, u) on the wire. Following from this, the kinetic energy of the bead at a time t after it is dropped is equal to

$$T = \frac{1}{2} m \left(\frac{ds}{dt}\right)^2,$$

while the potential energy is equal to

$$V = -mgu,$$

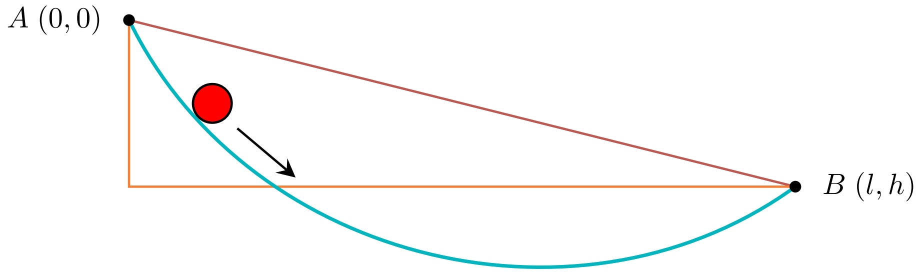

Figure 2. The brachistochrone setup, with possible paths.

decreasing from 0. Initially, both kinetic and potential energy are equal to 0. Energy is conserved so $T+V=0$ and we solve for ds/dt to obtain

$$\frac{ds}{dt}=\sqrt{2gu}.$$

Meanwhile, by the Pythagorean theorem we can obtain

$$ds=dx\sqrt{1+u'^2}$$

for the arclength element. As

$$dt=\frac{ds}{\sqrt{2gu}}=\frac{\sqrt{1+u'^2}}{\sqrt{2gu}}\,dx,$$

we take the integral of both sides to obtain

$$t=\frac{1}{\sqrt{2g}}\int_0^l\sqrt{\frac{1+u'^2}{u}}\,dx$$

for the total time taken by the bead. We minimize this by applying the Euler–Lagrange equation to the Lagrangian $L(x,u,u')=\sqrt{\frac{1+u'^2}{u}}$, giving (after considerable computation)

$$-\frac{\sqrt{1+u'^2}}{2u^{3/2}}-\frac{d}{dx}\frac{u'}{\sqrt{u(1+u'^2)}}=-\frac{2uu''+u'^2+1}{2\sqrt{u(1+u'^2)}}=0,$$

so we must solve the second order ordinary differential equation

$$2uu''+u'^2+1.$$

To do so, we use a powerful theorem:

Theorem 6.1. *If the Lagrangian $L(x,u,u')=L(u,u')$ does not depend on x, then the Hamiltonian function*

$$H(u,u')=L(u,u')-u'\frac{\partial L}{\partial \dot{u}}(u,u')$$

is a first integral *for the Euler–Lagrange equation, meaning that it is constant for each solution. In short,*

$$H(u(x),u'(x))=c$$

for some $c\in\mathbb{R}$ (which can depend upon the solution $u(x)$).

Proof. Differentiating the definition of the Hamiltonian, we obtain

$$\frac{d}{dx}H(u,u')=\frac{d}{dx}\left[L(u,u')-u'\frac{\partial L}{\partial \dot{u}}(u,u')\right]=u'\left(\frac{\partial L}{\partial u}(u,u')-\frac{d}{dx}\frac{\partial L}{\partial \dot{u}}(u,u')\right)=0$$

which equals zero directly by the Euler–Lagrange equation. Thus the Hamiltonian function is constant. ∎

The advantage of the equation $H(u(x), u'(x)) = c$ is that it is an implicit first order ODE which allows us to solve it by integration: reexpressing this as

$$u' = \frac{du}{dx} = h(u, c),$$

we can quickly obtain

$$\int \frac{du}{h(u,c)} = x + C$$

with constant C, which gives the general solution. For our brachistochrone example, the Hamiltonian function is

$$H(u, u') = L - u'\frac{\partial L}{\partial \dot{u}} = \frac{1}{\sqrt{u(1+u'^2)}} = c,$$

which we can rewrite to give

$$u(1 + u'^2) = k$$

for constant $k = 1/c^2$. We solve for u' and get

$$\frac{du}{dx} = \sqrt{\frac{k-u}{u}},$$

which we can easily solve via separation of variables to obtain

$$\int \sqrt{\frac{u}{k-u}}\, du = x + C$$

for some constant C. We integrate the left hand side with the trigonometric substitution $u = \frac{1}{2}k(1-\cos\theta)$, giving

$$x + C = \frac{k}{2}\int \sqrt{\frac{1-\cos\theta}{1+\cos\theta}} \sin\theta \, d\theta = \frac{k}{2}\int (1-\cos\theta)\, d\theta = \frac{1}{2}k(\theta - \sin\theta).$$

The left hand boundary condition implies that $C = 0$, so our solution to the Euler–Lagrange equation is the set of curves parametrized by

$$x = r(\theta - \sin\theta), \quad u = r(1-\cos\theta).$$

This curve is known as a *cycloid*, which is the curve traced by a point on the edge of a rolling circle. For our purposes, the right-hand boundary condition gives a unique value of r, and it is easy to see that this final curve indeed has the minimum time of descent.

7. Minimal Surfaces of Revolution

Another common application of the calculus of variations is in finding a surface with minimum area, satisfying certain constraints. These surfaces are often found in nature as soap films, which have been the object of fascination and study for centuries. In one general form, we have a simple closed curve $C \subset \mathbb{R}^3$ and look for the surface with the least total area out of all surfaces with C as their boundary. Thus, we seek to minimize the surface area integral $\iint_S dS$ over all surfaces $S \subset \mathbb{R}^3$ with the boundary curve $\partial S = C$. We call an area-minimizing surface like this a *minimal surface*.

In the general case, we can begin by assuming that the bounding curve C projects down to a simple closed curve $\Gamma = \partial\Omega$ that bounds an open domain $\Omega \subset \mathbb{R}^2$ in the (x, y) plane. We can then express the space curve $C \subset \mathbb{R}^3$ as $z = g(x, y)$ for $(x, y) \in \Gamma = \partial\Omega$. Generally we expect that the minimal surface S can be described as the graph of a function $z = u(x, y)$ parametrized by $(x, y) \in \Omega$. Thus the surface area is given by

$$J[u] = \iint_\Omega \sqrt{1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}\, dx\, dy.$$

We then want the function $z = u(x, y)$ that minimizes the surface area integral, following the boundary conditions

$$u(x, y) = g(x, y) \quad \text{for} \quad (x, y) \in \partial\Omega$$

that give the boundary curve C. However, in general the solutions to this are a complicated second order PDE.

For a more easily solvable case, we consider minimal *surfaces of revolution*, i.e. surfaces with rotational symmetry. Specifically, given two points (a, α) and (b, β), we attempt to find the curve $u(x)$ between them such that the surface formed by revolving it about the x-axis has the lowest possible surface area. Physically, this can be shown by stretching a soap film between two wire circles of radii α and β, a distance $b - a$ apart. Since the cross-sections of this surface of revolution are circles centered at the

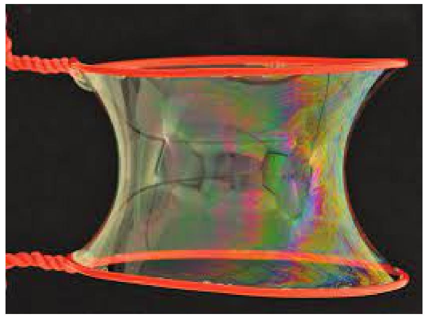

Figure 3. An example of a soap film that minimizes surface area.

x-axis, we can write the surface area as an integral

$$J[u] = \int_a^b 2\pi u\sqrt{1 + u'^2}\, dx$$

that we seek to minimize, given fixed boundary conditions $u(a) = \alpha, u(b) = \beta$. We also assume that the curve is given by a nonnegative function $y = u(x) \geq 0$. Simplifying this, we find the minima of the functional

$$J[u] = \int_a^b u\sqrt{1 + u'^2}\, dx,$$

with Lagrangian $L(x, u, u') = u\sqrt{1 + u'^2}$. We can find

$$\frac{\partial L}{\partial u} = \sqrt{1 + u'^2}, \quad \frac{\partial L}{\partial \dot{u}} = \frac{uu'}{\sqrt{1 + u'^2}}$$

so the Euler-Lagrange equation gives

$$\sqrt{1 + u'^2} - \frac{d}{dx}\frac{uu'}{\sqrt{1 + u'^2}} = \frac{1 + u'^2 - uu''}{(1 + u'^2)^{3/2}} = 0.$$

Since the Lagrangian does not depend upon x, we can use the Hamiltonian function to simplify this, giving

$$H(u, u') = L - u'\frac{\partial L}{\partial \dot{u}} = \frac{u}{\sqrt{1 + u'^2}} = c$$

for some constant c. Rearranging this to solve for u', we obtain

$$\frac{du}{dx} = u' = \frac{\sqrt{u^2 - c^2}}{c}$$

which we solve via separation of variables to give

$$\int \frac{c\ du}{\sqrt{u^2 - c^2}} = x + d,$$

with constant d. The left hand integral happens to be the inverse of the hyperbolic cosine function $\cosh x = \frac{1}{2}\left(e^x + e^{-x}\right)$, so

$$\cosh^{-1}\frac{u}{c} = x + d,$$

and we have the final equation

$$u(x) = c\ \cosh\left(\frac{x + d}{c}\right).$$

Any values of c and d that satisfy the boundary conditions will work. This curve in general is known as a catenary, and it is commonly found as the shape taken by hanging cables as well as being a strong arch for engineering.

8. Further Extensions

Here we give an overview of more advanced aspects in the calculus of variations, which go beyond the fundamentals we have explored.

One extension of basic functional calculus is studying the mathematics of different types of boundary conditions, as opposed to the fixed boundary conditions in the problems so far. Suppose we have a *free boundary*, with no boundary conditions on one or both ends of the interval. Here we can initially proceed similarly to how we evaluated the first variation with fixed boundaries, but we soon find that the boundary terms do not vanish, leaving us with

$$0 = v(b)\, \frac{\partial L}{\partial \dot{u}}\bigg|_{x=b} + \int_a^b \left(\frac{\partial L}{\partial u} - \frac{d}{dx}\frac{\partial L}{\partial \dot{u}} \right) v\, dx.$$

Since $v(b)$ may be anything, we can set it to 0 and thus imply that the integral term must still vanish. However, this means that for the whole expression to equal 0 for any $v(b)$, the term multiplying $v(b)$ must also equal zero. Thus we have the *natural boundary condition*

$$\frac{\partial L}{\partial \dot{u}}\,(b, u(b), u'(b)) = 0.$$

Similarly if the starting endpoint is not fixed, then it must satisfy

$$\frac{\partial L}{\partial \dot{u}}\,(a, u(a), u'(a)) = 0.$$

It is important to note that if another type of boundary condition — not fixed or natural — is imposed at an endpoint, then by the reasoning above the natural boundary condition must still be satisfied there. These conditions generally will prescribe a set value of u and u' there, meaning that only one solution exists, which typically does not satisfy the conditions at the other endpoint. In this case, there are no potential minimizing functions found through variations. For an example in practice, consider the brachistochrone problem from before, but now without a fixed endpoint: the horizontal distance remains the same, but the bead can end anywhere on a vertical line. Since the Lagrangian is

$$L(x, u, u') = \sqrt{\frac{1+u'^2}{u}} \qquad \text{with} \qquad \frac{\partial L}{\partial \dot{u}} = \frac{u'}{\sqrt{u(1+u'^2)}},$$

the natural boundary condition reduces to simply $u'(b) = 0$. Thus, the solution is the cycloid starting at the origin and with a horizontal tangent at $x = b$.

Another more complex development is the second variation, studying variations in a functional. This is analogous to the second derivative in standard calculus, and much like the second derivative test we can use the second variation to characterize critical points as maxima or minima. Recall that in multivariable calculus, the second partial derivative test uses the Hessian matrix, which collects all second partial derivatives of a function. For the second variation of a functional, we compute an integral of various second order derivatives of the Lagrangian — this is fairly convoluted and beyond the scope of this paper, but elaborated on in more detailed sources.

We can also continue by analyzing multiple dimensions, beyond a simple definite integral. In this case, the algebra is actually not so different, taking the derivative of a functional that consists of a multiple integral (like $J[u] = \iint_\Omega L(x, y, u, u_x, u_y)\, dx\, dy$) and deriving a higher-dimension Euler–Lagrange equation, expressed in terms of the total derivative. The calculus of variations is highly extensible, being a powerful analytical method to find solutions to many problems.

References

[1] John C Clegg. "Calculus of Variations (University Mathematical Texts Series)". In: ().
[2] J Figueroa-O'Farril. "Brief notes on the Calculus of Variations". In: (2016). URL: http://www.maths.ed.ac.uk/jmf/Teaching/Lectures/CoV.pdf.
[3] Peter J Olver. "The calculus of variations". In: *Applied Mathematics Lecture Notes. Sec* 21 (2012), p. 4. URL: https://www-users.cse.umn.edu/~olver/ln_/cvc.pdf.