An Introduction to the Calculus of Variations

Damon Falck

October 2017

Abstract

This paper is intended as an accessible introduction to the fundamentals of the calculus of variations, a sort of generalisation of calculus. After covering some basic prerequisites in multivariable calculus, I'll explain the derivation of the most important results in the field and then demonstrate their application to a few interesting and historic problems. The reader is assumed to have fluency in single-variable calculus including integration by parts.

Contents

1 Introduction

So far, we've come across the concept of optimising (finding maxima and minima of) functions in one or more dimensions with respect to some variable. However, often in mathematics and physics there comes the need to find some sort of *path* which either maximises or minimises a value — to find the best possible *function* for a given situation. This is the main task with which the calculus of variations is concerned.

Originally, the calculus of variations was Euler's response to the Brachistochrone problem posed by Johann Bernoulli in 1696, which I discuss later on in section 7. Lagrange and others then took this and developed it further into a rich field of interesting and useful mathematics.

Among the problems we can solve with the calculus of variations are:

- Finding the curve between two points that an object will slide down most quickly
- Finding the shape of a given perimeter that encloses the largest area
- Finding the shortest path between two points over some surface
- Finding the best shape of a container to retain heat, subject to certain constraints
- Many other problems

Most problems that involve finding some sort of optimal path or shape can be solved using the calculus of variations.

In this paper I start by touching on a couple of useful concepts from multivariate calculus, before going on to derive the Euler-Lagrange equation, the cornerstone of the calculus of variations. The actual derivation is quite dense and may be skipped by the reader. I'll then show how a few of the historic problems with which variational calculus is concerned can be solved.

Although the calculus of variations is normally a second-year option, its basics are actually easy to understand without much prior knowledge. I've tried to make this as accessible as I can.

2 Some prerequisites

While I'm assuming a comfortable knowledge of single-variable calculus, I just want to quickly mention a couple of key concepts from multivariable calculus that I'll use later on. I won't prove them here: this is just for reference.

2.1 Partial and total derivatives

The first is the partial derivative. Given a function $f(x, y, z)$, the partial derivative $\dfrac{\partial f}{\partial x}$ is defined as the ratio of the change in f to the change in x if we change x very slightly *while* keeping y and z constant. It is calculated just like a regular derivative.

The reason we use a special symbol is there is a distinction between this and the *total* derivative of a multivariate function. The total derivative of f with respect to x is written, familiarly, as $\dfrac{\mathrm{d}f}{\mathrm{d}x}$. If x, y and z are all independent then they're the same:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\partial f}{\partial x}\,.$$

However, if y and/or z have some *implicit* — that is, not explicitly written — dependency on x (for example, if they're both parameterised by the same variable), then the two are separate: changing x may result in a change in y and/or z as well. The total derivative takes this into account, while the partial derivative does not. This distinction will come into use later on.

2.2 The multivariate chain rule

The second thing I want to mention is how we extend the multivariate chain rule to multiple dimensions. If we have some function $f(x, y)$ but we parameterise x and y using some third variable t, so we write $x(t)$ and $y(t)$, then f is really just a function of t. We can then calculate its derivative as follows:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\,.$$

Intuitively this makes sense: the amount that f changes is made up of the change due to the change in x and the change due to the change in y. In general the partial derivative of a multivariate function with respect to a particular variable is just the sum of the partial derivatives of all inputs to that function with respect to that variable.

3 The shortest path between two points

Let's start by taking a look at one of the fundamental problems with which the calculus of variations is concerned. Say we have two points, A and B, with coordinates (x_1, y_1) and (x_2, y_2) respectively (see fig. 1). The task is to find the shortest path between A and B. This can be *any smooth curve* that passes through A and B — it could go to Saturn and back (but we're interested in the one with the shortest length).

At this point your intuition should probably be crying out in despair: *surely it's just a straight line between them!* It seems completely obvious that to get from A to B as quickly as possible, you need to go 'as the crow flies' — that is, in a straight line — but to actually prove this, as we'll soon find out, is decidedly non-trivial. Just think about how many ways you could get between the two points: how is it possible to show that a straight line is better than *absolutely everything else*?

To begin with, we need to formalise this problem algebraically. We're going to limit the possible curves to well-defined smooth functions[1] $y(x)$ between A and B. In order for $y(x)$ to pass

[1] The fact that we can do this is actually not immediately obvious, but we'll assume we can so as to make our lives easier.

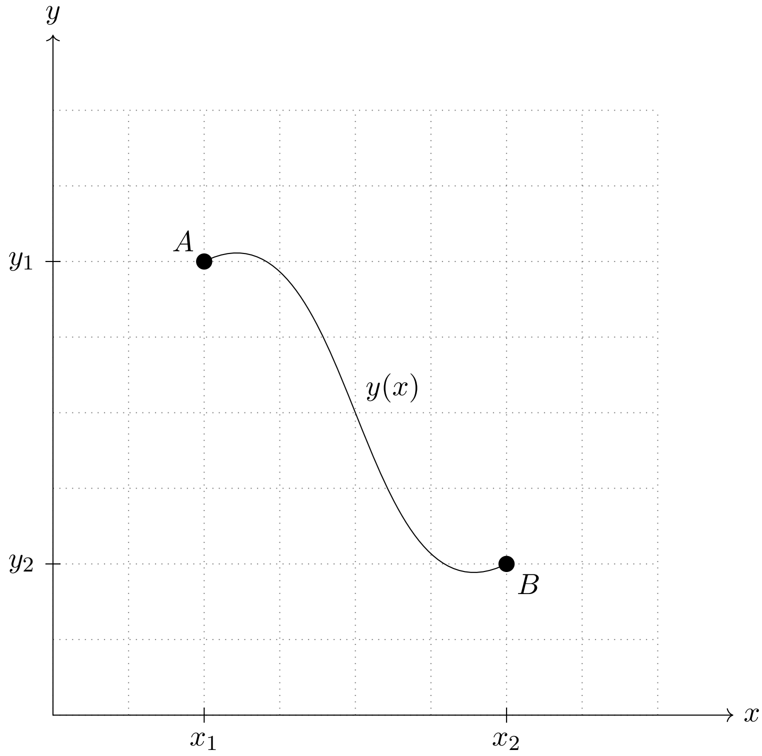

Figure 1: We want to find the shortest path between A and B.

through both points, we must have the constraints

$$y(x_1) = y_1 \quad \text{and} \quad y(x_2) = y_2.$$

Now take a small part of the curve, as shown in fig. 2, and call its length $\mathrm{d}s$. As $\mathrm{d}s$ approaches zero in length, it becomes a straight line, and so if the height of that section of the curve is $\mathrm{d}y$ and its width is $\mathrm{d}x$, then we can apply Pythagoras' theorem to say

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2.$$

Taking the square root and pulling out a factor of $\mathrm{d}x$ then brings us to

$$\begin{aligned} \mathrm{d}s &= \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} \\ &= \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\, \mathrm{d}x \end{aligned}$$

but the derivative $\frac{\mathrm{d}y}{\mathrm{d}x}$ can be written in terms of our function $y(x)$:

$$\mathrm{d}s = \sqrt{1 + y'^2(x)}\, \mathrm{d}x.$$

Now, if we want the total length of the whole curve $y(x)$ between $x = x_1$ and $x = x_2$ we just want to add up all these infinitesimal lengths $\mathrm{d}s$, meaning we take the integral:

$$L = \int_{x_1}^{x_2} \sqrt{1 + y'^2(x)}\, \mathrm{d}x.$$

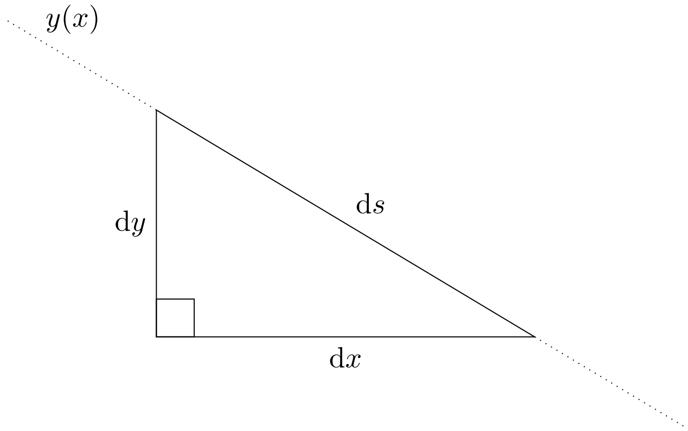

Figure 2: An infinitesimal section of the curve.

So, this total length L is what we want to minimise. Something strikes us as different, however, to minimisation problems we've come across before. Normally we just take the derivative of a function and set it to zero, but here L is dependent upon *every value of* the function $y(x)$ on the interval $[x_1, x_2]$. So L isn't a normal function: it takes a *function* as its input and it outputs a real number, and so we have to somehow differentiate with respect to all possible functions between A and B to find the function $y(x)$ that minimises L. The next section discusses just how we do that.

4 Introducing functionals and their variations

In the previous section we met a new type of function which is dependent upon *every* value of another function in a given interval. In fact, we call such a construct a *functional*.

Functionals can be thought of as 'functions of functions'. We use square brackets (as opposed to parentheses) to enclose their arguments, which are not numbers but functions themselves. For the total arc length we just derived, for example, we would write

$$L[y(x)] = \int_{x_1}^{x_2} \sqrt{1 + y'^2(x)}\,\mathrm{d}x.$$

In fact, most functionals can be expressed as a definite integral of some combination of a function, its arguments and its derivatives: let's consider a general functional $F[y(x)]$ defined as[2]

$$F[y] = \int_{x_1}^{x_2} f(x, y(x), y'(x))\,\mathrm{d}x \tag{1}$$

for some real-valued smooth function f. We keep our constraints that $y(x_1) = y_1$ and $y(x_2) = y_2$.

Our objective is to find the extrema of this functional F — that is, we want to find which function $y(x)$ either maximises or minimises $F[y]$. Just like in regular calculus, this is going to

[2]Note that writing $F[y]$ is entirely equivalent to writing $F[y(x)]$, as most often the actual number x isn't relevant; in the integral it's just the 'dummy variable'.

involve some sort of derivative, but in our case we want to know how much the real number $F[y]$ will vary as a result of a small change in shape of the function y. When this variation is zero, just like in normal calculus, we have found an extremum of F (a stationary point).

Consider our function $y(x)$ and suppose we make a tiny change to its shape by adding to it a small amount ε of some arbitrary smooth function $\eta(x)$. Note that we require the resulting function still to pass through the points (x_1, y_1) and (x_2, y_2) (for instance in the shortest path problem these are the two points our curve *must* connect) and so at $x = x_1$ and $x = x_2$ this new function $\eta(x)$ must not make any difference to the value of y — that is, we require

$$\eta(x_1) = \eta(x_2) = 0.$$

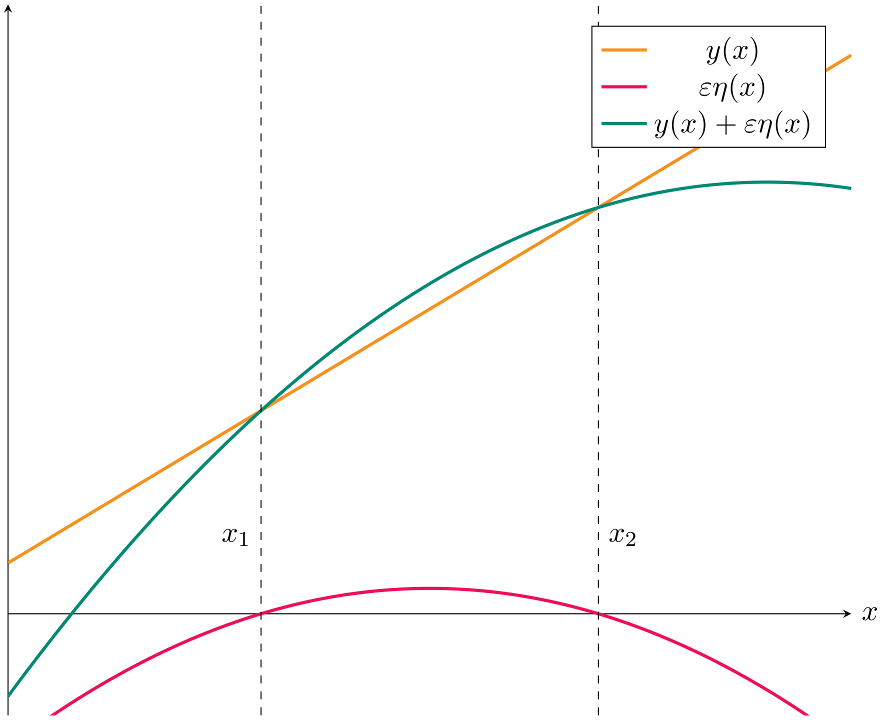

Figure 3: An example of what the functions $y(x)$, $\eta(x)$ and $y(x) + \varepsilon\eta(x)$ might look like.

Then, the difference between the value of the functional when applied to this new, slightly different function $y(x) + \varepsilon\eta(x)$, and its value when applied to $y(x)$, all divided by that amount ε, is defined as the functional's *variation* or *first variation*:

Definition. The variation of a functional $F[y]$ *in the direction of* an arbitrary function η is the ratio of the change in F to the change in ε when the shape of its input function y is changed by an infinitesimal amount ε of η, and is defined as

$$\delta F(y; \eta) \coloneqq \lim_{\varepsilon \to 0} \frac{F[y(x) + \varepsilon\eta(x)] - F[y(x)]}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{F[y + \varepsilon\eta] - F[y]}{\varepsilon}. \tag{2}$$

Compare this with the definition of a regular derivative and there is striking similarity; what we're doing is conceptually very closely related. We're changing the input of something by an infinitesimal amount and seeing how much the output changes as a result.[3]

We're next going to try to come up with a more useful form for the variation of a functional. Consider the functional $F[y+\varepsilon\eta]$ (as opposed to just $F[y]$ as before). By comparison[4] with eq. (1),

$$F[y+\varepsilon\eta] = \int_{x_1}^{x_2} f(x, y(x)+\varepsilon\eta(x), y'(x)+\varepsilon\eta'(x))\,\mathrm{d}x.$$

If you imagine temporarily that the shapes of $y(x)$ and $\eta(x)$ are fixed but we can freely change the value of ε, then F is just a function of ε:

$$F(\varepsilon) = \int_{x_1}^{x_2} f(x, y(x)+\varepsilon\eta(x), y'(x)+\varepsilon\eta'(x))\,\mathrm{d}x.$$

Differentiating with respect to ε (and using our definition of the derivative) gives

$$\frac{\mathrm{d}F}{\mathrm{d}\varepsilon} = \lim_{h\to 0}\frac{F(\varepsilon+h)-F(\varepsilon)}{h}$$

which is just the same as writing

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}F[y+\varepsilon\eta] = \lim_{h\to 0}\frac{F[y+(\varepsilon+h)\eta]-F[y+\varepsilon\eta]}{h}.$$

So, evaluating this derivative at $\varepsilon = 0$ gives

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}F[y+\varepsilon\eta]\right|_{\varepsilon=0} = \lim_{h\to 0}\frac{F[y+h\eta]-F[y]}{h}.$$

But, replace[5] every h with an ε and you see from eq. (2) that this is *exactly the same* as our definition of the variation of $F[y]$ in the direction of η! So, our more useful definition is:

Definition. The variation of a functional $F[y]$ in the direction of an arbitrary function η is given by

$$\delta F(y;\eta) = \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}F[y+\varepsilon\eta]\right|_{\varepsilon=0}. \tag{3}$$

It is this form that will allow us to derive the all-important Euler-Lagrange equation in the next section.

5 Deriving the Euler-Lagrange equation

Now that we have a nice expression for the variation of a functional $F[y]$, we return to our original problem which is to find the function y that maximises or minimises F. Just like we

[3]For those of you who have studied multivariate calculus, this is actually really just a directional derivative, except here our direction is not a vector but a function. If you like, it's an infinite-dimensional vector.

[4]To find the third argument to f we just had to differentiate $y(x)+\varepsilon\eta(x)$ with respect to x, resulting in $y'(x)+\varepsilon\eta'(x)$.

[5]We can do this as h is just a dummy variable inside the limit.

would set the derivative of a function to zero to find its extrema, we set the variation of a functional to zero to do the same:

$$\begin{aligned} \delta F(y;\eta) &= 0 \\ \implies \frac{\mathrm{d}}{\mathrm{d}\varepsilon} F[y+\varepsilon\eta]\bigg|_{\varepsilon=0} &= 0 \qquad \text{(from eq. (3))}. \end{aligned} \tag{4}$$

In fact, in this situation our expression for the variation makes even more intuitive sense. Suppose y *is indeed* the function which minimises F; then, changing its shape in any way — that is, adding any amount ε of any other function $\eta(x)$ — will surely increase F. That is, the minimum of $F[y+\varepsilon\eta]$ is at $\varepsilon = 0$. Therefore regular calculus tells us that the derivative with respect to ε at $\varepsilon = 0$ *must* be zero — it's a stationary point. This argument is unchanged if y maximises F instead, and either way leads to eq. (4).

So now our task is to somehow use eq. (4) to solve for the function y that minimises or maximises F.

Let's use our definition of F from eq. (1). We clearly require

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_{x_1}^{x_2} f(x, y(x)+\varepsilon\eta(x), y'(x)+\varepsilon\eta'(x))\,\mathrm{d}x\bigg|_{\varepsilon=0} = 0$$

and to make our lives easier, writing $Y(x) = y(x) + \varepsilon\eta(x)$, this becomes

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_{x_1}^{x_2} f(x, Y(x), Y'(x))\,\mathrm{d}x\bigg|_{\varepsilon=0} = 0$$

Using the fact that differentiation and integration of smooth functions can be done in any order[6], we can rewrite this condition as[7]

$$\int_{x_1}^{x_2} \frac{\partial}{\partial\varepsilon} f(x, Y(x), Y'(x))\bigg|_{\varepsilon=0}\,\mathrm{d}x = 0. \tag{5}$$

Next, to attack this further, we'll apply the multivariate chain rule we discussed in section 2, which gives

$$\frac{\partial f}{\partial\varepsilon} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial\varepsilon} + \frac{\partial f}{\partial Y}\frac{\partial Y}{\partial\varepsilon} + \frac{\partial f}{\partial Y'}\frac{\partial Y'}{\partial\varepsilon}\,.$$

Of course x doesn't depend on ε at all, so $\frac{\partial x}{\partial\varepsilon} = 0$ meaning we can get rid of the first term:

$$\frac{\partial f}{\partial\varepsilon} = \frac{\partial f}{\partial Y}\frac{\partial Y}{\partial\varepsilon} + \frac{\partial f}{\partial Y'}\frac{\partial Y'}{\partial\varepsilon}\,.$$

Now it can be seen that

$$\frac{\partial Y}{\partial\varepsilon} = \frac{\partial}{\partial\varepsilon}\big(y(x)+\varepsilon\eta(x)\big) = \eta(x)$$

[6]This is non-trivial to prove so we'll just take it as an assumption

[7]The reason that we now use a partial derivative is that the only variable on which the value of the whole integral depends is ε and so we can take an ordinary derivative with respect to ε; but the actual function f *inside* the integral depends on x as well and so if we bring our derivative inside the integral we *must* make it a partial derivative to make clear that we're holding x constant and differentiating with respect to ε only.

and similarly

$$\frac{\partial Y'}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon}\left(y'(x) + \varepsilon\eta'(x)\right) = \eta'(x).$$

Thus, our partial derivative of f with respect to ε is simply

$$\frac{\partial f}{\partial \varepsilon} = \frac{\partial f}{\partial Y}\,\eta(x) + \frac{\partial f}{\partial Y'}\,\eta'(x),$$

making our optimality condition from eq. (5) become

$$\int_{x_1}^{x_2} \left(\frac{\partial f}{\partial Y}\,\eta(x) + \frac{\partial f}{\partial Y'}\,\eta'(x)\right)\Bigg|_{\varepsilon=0} \mathrm{d}x = 0.$$

Remembering that we defined $Y(x)$ as equal to $y(x)+\varepsilon\eta(x)$, at $\varepsilon = 0$ we must have $Y(x) = y(x)$ (and similarly $Y'(x) = y'(x)$), so we can make this replacement and split it into two integrals:

$$\int_{x_1}^{x_1} \frac{\partial f}{\partial y}\,\eta(x)\,\mathrm{d}x + \int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\,\eta'(x)\,\mathrm{d}x = 0. \tag{6}$$

Remember here that we're eventually trying to get rid of any mention of the function η as it is by definition just some arbitrary function. The right-hand integral above can be evaluated using integration by parts. With reference to the parts formula,

$$\int u'v\,\mathrm{d}x = uv - \int uv'\,\mathrm{d}x,$$

we set $u' = \eta'(x)$ so that $u = \eta(x)$, and so our other function is $v = \dfrac{\partial f}{\partial y'}$ meaning $v' = \dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{\partial f}{\partial y'}\right)$. Then,

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\,\eta'(x)\,\mathrm{d}x = \left[\frac{\partial f}{\partial y'}\,\eta(x)\right]_{x_1}^{x_2} - \int_{x_1}^{x_2} \eta(x)\,\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)\mathrm{d}x.$$

How has this helped us? Well, you may remember that when we introduced the function η, we made sure to mention that it has to be constrained by $\eta(x_1) = \eta(x_2) = 0$ (see fig. 3). Therefore,

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

and so coming back to eq. (6), we now need

$$\int_{x_1}^{x_1} \frac{\partial f}{\partial y}\,\eta(x)\,\mathrm{d}x - \int_{x_1}^{x_2} \eta(x)\,\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)\mathrm{d}x = 0.$$

Combining the integrals and factoring out $\eta(x)$ leaves us with

$$\int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)\right)\eta(x)\,\mathrm{d}x = 0.$$

We're extremely close now. Since η is arbitrary — that is, it could be *any* function as long as $\eta(x_1) = \eta(x_2) = 0$ — we can conclude that the only way this integral is *always* zero, for *any* function η, is if the other function in the integrand is always zero, that is,

$$\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = 0 \tag{7}$$

for every value of x on the interval $[x_1, x_2]$. This reasoning is slightly opaque and in fact the argument is known appropriately as 'the fundamental lemma of the calculus of variations'. If you're interested, here's a short formal proof:

Lemma (The Fundamental Lemma of the Calculus of Variations)**.** *Let $\phi(x)$ be a continuous function on the interval $[x_1, x_2]$. Suppose that for all continuous functions $\eta(x)$ with $\eta(x_1) = \eta(x_2) = 0$ we have*

$$\int_{x_1}^{x_2} \phi(x)\eta(x)\,\mathrm{d}x = 0. \tag{8}$$

Then $\phi(x) = 0$ for all $x \in [x_1, x_2]$.

Proof. Suppose that there is some value $\alpha \in [x_1, x_2]$ for which $\phi(\alpha) \neq 0$. Without loss of generality we may say that here $\phi(\alpha) > 0$. Then as ϕ is continuous, there is some interval $[\alpha_1, \alpha_2]$, where $\alpha_1 \leqslant \alpha \leqslant \alpha_2$, on which ϕ is always positive. Suppose η is positive on this interval and zero elsewhere; then

$$\int_{x_1}^{x_2} \phi(x)\eta(x)\,\mathrm{d}x = \int_{\alpha_1}^{\alpha_2} \phi(x)\eta(x)\,\mathrm{d}x > 0$$

which contradicts eq. (8). Hence there cannot exist any $\alpha \in [x_1, x_2]$ for which $\phi(\alpha) \neq 0$. □

Therefore, in eq. (7) we end up at a final, beautiful differential equation which is called the Euler-Lagrange equation. It guarantees the following:

Theorem (The Euler-Lagrange Equation)**.** *Any functional*

$$F[y] = \int_{x_1}^{x_2} f(x, y(x), y'(x))\,\mathrm{d}x$$

will take an extremal value if and only if the function y satisfies the differential equation

$$\frac{\partial f}{\partial y} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \tag{9}$$

for every $x \in [x_1, x_2]$.

This is exactly what we were hoping for: now given any functional we want to maximise or minimise, we need only solve this differential equation for the function y at which such a stationary point will occur.

6 Returning to the shortest path

In this section we return to the shortest path problem introduced in section 3; we now have all the tools to solve it. As we found out, the total length of a curve $y(x)$ between two points

(x_1, y_1) and (x_2, y_2) is given by

$$L[y] = \int_{x_1}^{x_2} \sqrt{1 + y'^2(x)}\,\mathrm{d}x,$$

and we want to use the Euler-Lagrange equation derived in section 5 to find the curve y which minimises this length. In this case, the function f referred to by the Euler-Lagrange equation is

$$f(y'(x)) = \sqrt{1 + y'^2(x)}.$$

The function f depends only on y', not y itself, and so $\frac{\partial f}{\partial y} = 0$. Differentiating with respect to y',

$$\begin{aligned} \frac{\partial f}{\partial y'} &= \frac{1}{2}(1 + y'^2)^{-\frac{1}{2}} \cdot 2y' \\ &= \frac{y'}{\sqrt{1 + y'^2}}, \end{aligned}$$

and so the Euler-Lagrange equation, eq. (9), gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y'}{\sqrt{1 + y'^2}}\right).$$

The derivative with respect to x is always zero — which is the same as saying there is some constant k such that for all x,

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

Rearranging this leads us to

$$\begin{aligned} y' &= k\sqrt{1 + y'^2} \\ \implies y'^2 &= k^2(1 + y'^2) \\ \implies y'^2 &= \frac{k^2}{1 - k^2} \\ \implies y' &= \frac{k}{\sqrt{1 - k^2}}. \end{aligned}$$

Therefore it can be seen that $y'(x)$, the derivative of the curve with respect to x, is a constant. In other words, that curve is a straight line.

Of course, it seems quite superfluous to have gone through so much abstract mathematics to get to this point, but there is real value in proving this rigorously. While our intuition is correct in this case, a lot of the time it may not be and we *must* resort to mathematics for the answer. For instance, we know that if these two points are on a flat plane, the shortest path between them is a straight line — but what if they're on a sphere, or a pseudosphere (a surface with outward curvature)? What route should an aeroplane take from London to New York to get there in the shortest time? These are questions which we need the calculus of variations to answer.

7 The brachistochrone problem

Brachistochrone literally means 'least time' in Greek. The problem is an historic one, and perhaps the most perfect example of a useful application of the calculus of variations. Indeed, it was Johann Bernoulli's raising of the problem in 1696 that eventually led to Euler's 1756 publication *Elementa Calculi Variationum* which kick-started the field.

We start with two points A and B in the same vertical plane. Imagine releasing a ball or other massive particle at point A. We want to find the smooth curve from A to B down which a particle will slide, without friction, in the shortest time possible. That is, we want to minimise the total time taken between the particle being released from rest at A and arriving at B.

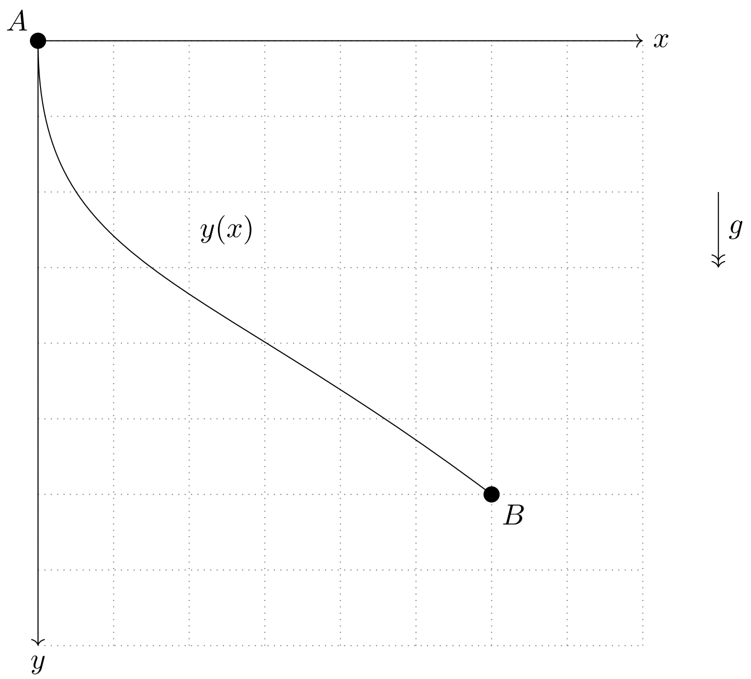

Figure 4: For this problem we want to find the 'brachistochrone' curve from A to B down which an object acting under gravity and without friction will slide in the least time.

Unlike the shortest path problem discussed in sections 3 and 6, the solution is not immediately obvious. Sliding down a straight line, it turns out, is surprisingly slow, and there are all sorts of other types of curves we could draw from A to B. We resort then to our newly acquired skills in the calculus of variations.[8]

[8] There are actually several other ways to solve this problem, perhaps the nicest using an analogy with optics. Fermat's principle states that light will always take the path of least time between two points, and so by shining a ray of light at an angle through a medium of constantly changing optical density (such that the light experiences a downward acceleration of g), we can use Snell's law of refraction to find what path the light will take. This solution was originally proposed by Johann Bernoulli in 1697, long before Euler's paper on the calculus of variations was published.

7.1 Setting up the integral

Let A be at the origin and let B have coordinates (x_0, y_0). We set our axes such that *downwards* is the positive y-direction; this will save us some hassle later on. Our task is to minimise the total time taken for the particle to slide from A to B, so we first want an expression for the total time in terms of things we know.

The particle is released from rest at A so has zero kinetic energy at this point. Let's also set the gravitational potential energy to be zero here, so that the particle has a total energy of zero. Then at some arbitrary height y below A, if the particle has speed v then it has kinetic energy $\frac{1}{2}mv^2$ and potential energy $-mgy$ (as its elevation has decreased) — but by energy conservation the total energy must still be zero, so

$$\begin{aligned} \frac{1}{2}mv^2 - mgy &= 0 \\ \implies \frac{1}{2}v^2 &= gy \\ \implies v &= \sqrt{2gy}. \end{aligned}$$

This is the *speed* of the particle in the direction of the curve. Since we don't know the length of the curve, we want to find either the vertical or horizontal velocity instead. The horizontal velocity of the particle at this point is given by

$$v_x = v \sin\theta = \sqrt{2gy}\sin\theta$$

where θ is the angle the curve makes to the vertical.

Now take a small length $\mathrm{d}s$ of the curve, as in fig. 5.

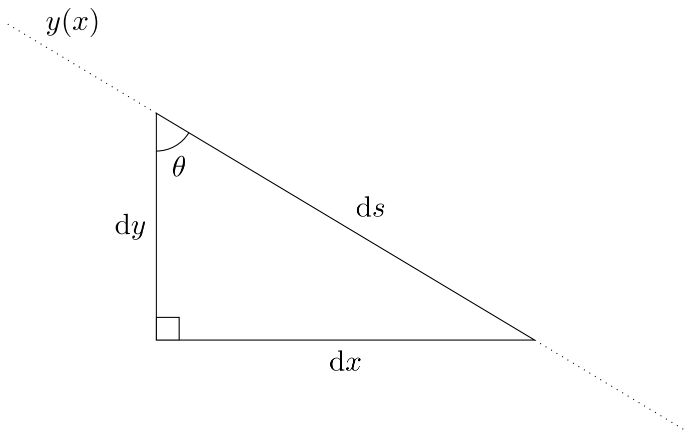

Figure 5: An infinitesimal section of the curve.

It's clear from this diagram that $\sin\theta = \frac{\mathrm{d}x}{\mathrm{d}s}$ and so we can simplify our expression for the horizontal velocity to

$$v_x = \sqrt{2gy}\,\frac{\mathrm{d}x}{\mathrm{d}s}\,.$$

but looking again at fig. 5, Pythagoras' theorem tells us that $\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}$ and so our expression becomes

$$\begin{aligned} v_x &= \sqrt{\frac{2gy(\mathrm{d}x)^2}{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}} \\ &= \sqrt{\frac{2gy(x)}{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}} \\ &= \sqrt{\frac{2gy}{1 + y'^2(x)}}. \end{aligned}$$

We now have a nice expression for the horizontal velocity at every point, and we know that the total horizontal distance the particle needs to travel is x_0, so by analogy with the linear formula

$$\text{time} = \frac{\text{displacement}}{\text{velocity}},$$

to find the total time taken T we take the definite integral:[9]

$$\begin{aligned} T &= \int_0^{x_0} \frac{\mathrm{d}x}{v_x} \\ \implies T[y] &= \int_0^{x_0} \sqrt{\frac{1 + y'^2}{2gy}}\, \mathrm{d}x. \end{aligned}$$

This is the functional we want to minimise over all functions $y(x)$. We know everything we need to know to do this now!

7.2 Solving using the Euler-Lagrange equation

The Euler-Lagrange differential equation we derived in section 5 tells us that for the function $y(x)$ to minimise T we must have

$$\frac{\partial f}{\partial y} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \tag{10}$$

where in this case the function f referred to is the integrand,

$$f(y, y') = \sqrt{\frac{1 + y'^2}{2gy}}.$$

We need therefore to work out $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial y'}$. The former is

$$\begin{aligned} \frac{\partial f}{\partial y} &= \frac{1}{2}\sqrt{\frac{2gy}{1 + y'^2}}\left(-\frac{1 + y'^2}{2gy^2}\right) \\ &= -\frac{\sqrt{2}}{4}\sqrt{\frac{1 + y'^2}{gy^3}}, \end{aligned} \tag{11}$$

[9] Another way to see that we have to take this integral is by looking at how much time the particle takes to traverse the small distance $\mathrm{d}x$ and adding up all these times.

and the latter is

$$\frac{\partial f}{\partial y'} = \frac{1}{2}\sqrt{\frac{2gy}{1+y'^2}}\left(\frac{y'}{gy}\right)$$
$$= \frac{\sqrt{2}}{2}\left(\frac{y'}{\sqrt{gy(1+y'^2)}}\right). \tag{12}$$

Looking again at eq. (10), we need to differentiate this with respect to x. Of course, we have simply

$$\frac{\mathrm{d}y}{\mathrm{d}x} = y' \qquad \text{and} \qquad \frac{\mathrm{d}y'}{\mathrm{d}x} = y'',$$

so using the quotient rule with reference to eq. (12) we come to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\sqrt{2}}{2}\left[\frac{y''\sqrt{gy(1+y'^2)} - y'\cdot\frac{1}{2\sqrt{gy(1+y'^2)}}\cdot\left(gy'(1+y'^2)+gy(2y'y'')\right)}{gy(1+y'^2)}\right].$$

This is particularly nasty, but we can simplify it down a bit:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\sqrt{2}}{2}\left[\frac{y''}{\sqrt{gy(1+y'^2)}} - \frac{y'^2}{2y^{\frac{3}{2}}\sqrt{g(1+y'^2)}} - \frac{y'^2y''}{\sqrt{gy}(1+y'^2)^{\frac{3}{2}}}\right].$$

So, using this and eq. (11), the Euler-Lagrange equation gives

$$-\frac{\sqrt{2}}{4}\sqrt{\frac{1+y'^2}{gy^3}} = \frac{\sqrt{2}}{2}\left[\frac{y''}{\sqrt{gy(1+y'^2)}} - \frac{y'^2}{2y^{\frac{3}{2}}\sqrt{g(1+y'^2)}} - \frac{y'^2y''}{\sqrt{gy}(1+y'^2)^{\frac{3}{2}}}\right]$$
$$\implies -\frac{1}{2}\sqrt{\frac{1+y'^2}{gy^3}} = \frac{y''}{\sqrt{gy(1+y'^2)}} - \frac{y'^2}{2y\sqrt{gy(1+y'^2)}} - \frac{y'^2y''}{(1+y'^2)\sqrt{gy(1+y'^2)}}.$$

Multiplying through by $\sqrt{gy(1+y'^2}$, we get

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

which simplifies down to

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

or, rearranging slightly,

$$y'^2 + 2yy'' + 1 = 0.$$

This is a second-order differential equation that we can solve to find the curve $y(x)$ which minimises the total time! Multiplying by y' as our integrating factor, we have

$$y'^3 + 2yy'y'' + y' = 0$$

and after staring at the left hand side for a short while while thinking about the product rule, we see that this is the same as saying

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(yy'^2 + y\right) = 0.$$

So,

$$yy'^2 + y = a$$

for some constant a. Rearranging slightly and re-writing y' as $\frac{\mathrm{d}y}{\mathrm{d}x}$, this becomes

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \frac{a-y}{y}. \tag{13}$$

We've reduced the problem to a first-order differential equation! Of course, we could slog through and find the general solution by separating the variables, and then use the constraint that the curve must pass through points A and B to find the specific solution. But, when Bernoulli arrived at this equation from his own method, he instantly recognised it as the equation of a *cycloid*.

7.3 Showing that the optimal curve is a cycloid

A cycloid is what we call the path that a point on the circumference of a circle traces out as the circle rolls along a flat surface. A standard such curve is shown in fig. 6.

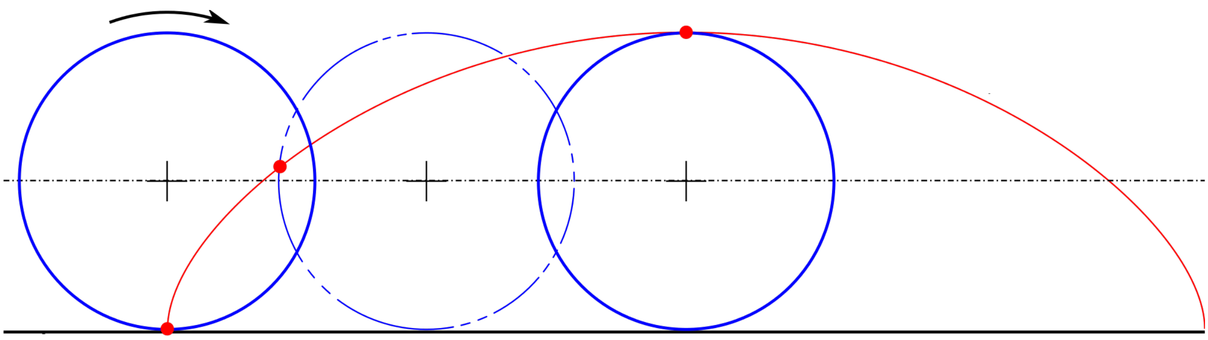

Figure 6: A cycloid curve traced out by a circle rolling along a horizontal surface. (Image source: Wikimedia Foundation.)

In the case of the brachistochrone, of course, the imaginary 'circle' that we roll to get our cycloid path will be rolling along the *ceiling* — the x-axis. It'll look something like fig. 7.

Let's see why such a cycloid gives the same differential equation as eq. (13). We want to find the square of the derivative of such a cycloid, so that we can directly compare the two. In fact, we'll just use Euclidean geometry for this part of the proof.

Point A is, as before, at the origin, as shown in fig. 8. Let our circle of radius r have centre C and be tangential to the x-axis at D. The point on the circumference of the circle that is drawing our cycloid curve is E and has coordinates (x, y). (Remember that y gets larger

Figure 7: The approximate shape of the cycloid path between A and B.

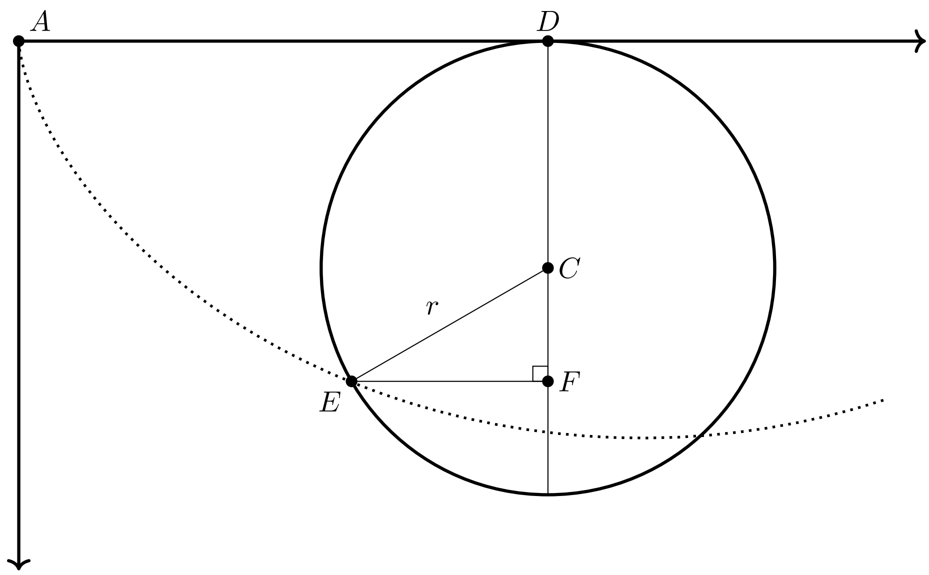

Figure 8: A snapshot of the circle in motion drawing our brachistochrone cycloid.

downwards.) Now drop a perpendicular from E to the vertical diameter of the circle and call this new point F.

The circle started with point E at the origin and rolled along the x-axis, so the total distance rolled is clearly

$$\overline{AD} = \overline{DE},$$

where $\overline{DE}$ means the arc length between D and E.

Therefore, the x-coordinate of point E is just

$$x = \overline{AD} - \overline{EF} = \overline{DE} - \overline{EF}, \tag{14}$$

and the y-coordinate is

$$y = \overline{DF}. \tag{15}$$

So, let's see what happens when the circle rotates very slightly and E is advanced to a new point G.

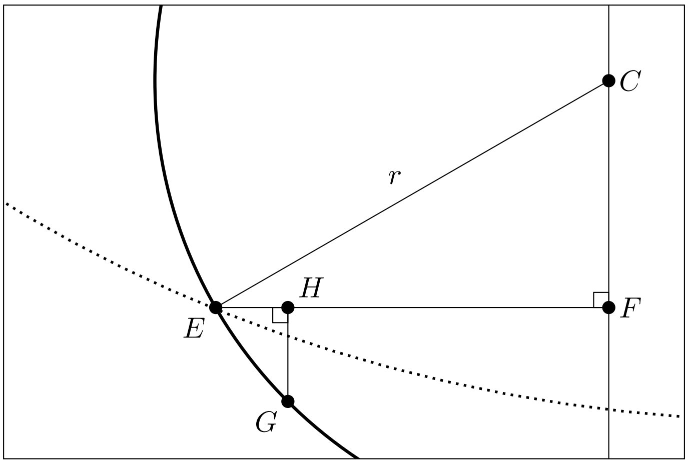

Figure 9: Point E is advanced slightly to G.

Using point H marked in fig. 9, the change in E's y-coordinate is

$$\mathrm{d}y = \overline{GH}$$

and the change to its x-coordinate is

$$\mathrm{d}x = \overline{EG} - (-\overline{EH})$$

by analogy with eqs. (14) and (15). So, we can say that the derivative of the cycloid path E makes is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{G \to E} \frac{\overline{GH}}{\overline{EG} + \overline{EH}}.$$

However, as G approaches E, the arc EG becomes a straight line and the triangle $\triangle EGH$ approaches similarity with $\triangle ECF$. So, our derivative becomes

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\overline{EF}}{\overline{CE} + \overline{CF}}.$$

We wanted to find the *square* of the derivative, so squaring this gives

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \frac{\overline{EF}^2}{(\overline{CE} + \overline{CF})^2}.$$

By Pythagoras, $\overline{EF}^2 = \overline{CE}^2 - \overline{CF}^2$ and so

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \frac{\overline{CE}^2 - \overline{CF}^2}{(\overline{CE} + \overline{CF})^2}.$$

Looking at this, though, the distance $\overline{CE}$ is just the radius r of the circle, and similarly

$$\overline{CF} = y - r.$$

So, finally,

$$\begin{aligned}
\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 &= \frac{r^2-(y-r)^2}{(r+y-r)^2} \\
&= \frac{r^2-y^2+2ry-r^2}{y^2} \\
&= \frac{2ry-y^2}{y^2} \\
&= \frac{2r-y}{y}.
\end{aligned}$$

Compare this to the differential equation for the brachistochrone we derived in eq. (13); if our arbitrary constant is $a = 2r$, they are exactly the same! Thus, the least-time curve between two points is a cycloid.

How do we find the *exact* cycloid that will get us from A to B in the shortest time? Just draw a cycloid starting from point A (the origin) as in fig. 7 and then scale it until it intersects point B. There's only one that will!

This is a wonderful proof and its consequences are rather unintuitive. As can be seen, depending on the positions of A and B the curve may actually slope upwards at the end towards B. Thinking about why this may occur helps: sometimes building greater speed by accelerating earlier will outweigh the slowing effect of the upwards climb at the end.

8 Proving the fundamental theorem of calculus

Another corollary hidden away in the powers of the Euler-Lagrange equation is a quick proof of the fundamental theorem of calculus for continuous functions! Suppose we want find the continuous function y through points (x_1, y_1) and (x_2, y_2) that minimises the functional

$$I = \int_{x_1}^{x_2} y'(x)\,\mathrm{d}x.$$

Applying the Euler-Lagrange equation, where now the function f it talks about is y', we get

$$\begin{aligned}
\frac{\partial f}{\partial y} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \\
\iff \frac{\mathrm{d}y'}{\mathrm{d}y} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial y'}{\partial y'}\right) \\
\iff 0 &= \frac{\mathrm{d}}{\mathrm{d}x}(1) \\
\iff 0 &= 0.
\end{aligned}$$

This tells us that y minimises $\int_{x_1}^{x_2} y'(x)\,\mathrm{d}x$ if and only if $0 = 0$. Since the latter is always true, *all* functions y minimise that functional. That is, the value of the definite integral $\int_{x_1}^{x_2} y'(x)\,\mathrm{d}x$ is the same for all functions y that pass through (x_1, y_1) and (x_2, y_2).

Suppose $y(x)$ is a straight line of equation

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1).$$

So, $y'(x)$ is

$$y' = \frac{y_2 - y_1}{x_2 - x_1},$$

a constant. Thus in this case, the definite integral of y' between x_1 and x_2 is just the area of a rectangle of width $x_2 - x_1$ and height y'. That is,

$$I = (x_2 - x_1)y' = (x_2 - x_1)\left(\frac{y_2 - y_1}{x_2 - x_1}\right) = y_2 - y_1.$$

But we just showed that the value of this integral is the same for *all* functions y, and so we have the general identity

$$\int_{x_1}^{x_2} y'(x)\,\mathrm{d}x = y_2 - y_1.$$

By definition $y_1 = y(x_1)$ and $y_2 = y(x_2)$, so we have

$$\int_{x_1}^{x_2} y'(x)\,\mathrm{d}x = y(x_2) - y(x_1).$$

Let $g(x) = y'(x)$ and $G(x) = y(x)$ be an antiderivative of $g(x)$. Then we have the first fundamental theorem of calculus in its usual form:

$$\int_{x_1}^{x_2} g(x)\,\mathrm{d}x = G(x_2) - G(x_1).$$

The second fundamental theorem of calculus, or the Newton-Leibniz axiom, is really just a generalisation of this to non-continuous functions.

9 The isoperimetric problem

Here's another classic problem the calculus of variations can be used to solve. We want to find the plane figure of a fixed perimeter ('isoperimetric') that encloses the largest possible area. The Greeks correctly convinced themselves that this was a circle, but what follows is a rigorous proof.

Most textbooks begin by taking a portion of the curve between $(0, 0)$ and $(1, 1)$ and finding the Cartesian equation of the curve between the two points that maximises the area, given a fixed arc length.

Instead, I'm going to use polar coordinates as I think this way the result will be much nicer. Let our shape be given by the polar function $r = f(\theta)$. (Here, r is the distance from the origin and θ is the angle made anticlockwise from the positive x-axis.)

Let's start by finding an expression for the area of the shape. Take a tiny wedge from the shape as in fig. 11.

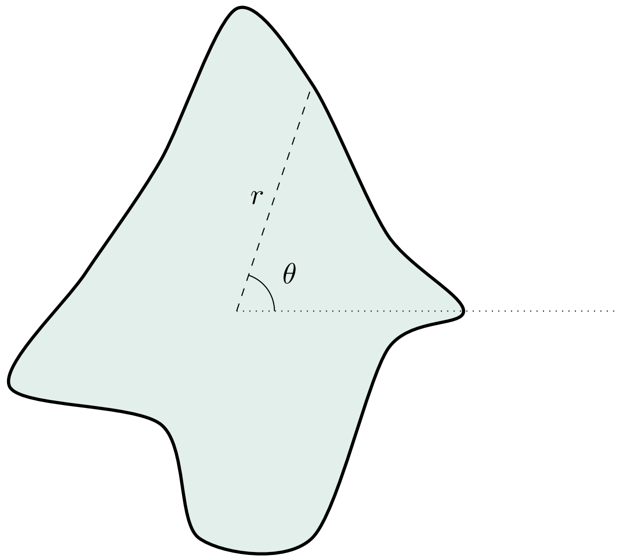

Figure 10: A generalised, blobby, polar curve.

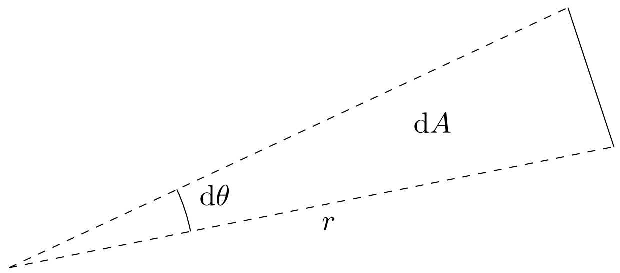

Figure 11: A small portion of the area enclosed by a polar curve.

We can actually approximate the area of this wedge by imagining it's a sector of a circle. Then, it has area $\frac{1}{2}r^2\,\mathrm{d}\theta$. Summing over all of these wedges therefore, the total area enclosed by the shape is

$$A = \int_0^{2\pi} \frac{1}{2}r^2\,\mathrm{d}\theta = \int_0^{2\pi} \frac{1}{2}f^2(\theta)\,\mathrm{d}\theta.$$

Looking at the same wedge, and still approximating it as a sector of a circle, the circular arc has length $r\,\mathrm{d}\theta$ and so the perimeter of the entire shape is

$$P = \int_0^{2\pi} r\,\mathrm{d}\theta = \int_0^{2\pi} f(\theta)\,\mathrm{d}\theta.$$

Let's fix this perimeter as l. We see that this is a different kind of problem to previously: we want to minimise the functional $A[f]$ *subject to the constraint* that $P[f] = l$.

Here we resort to using a Lagrange multipliers, a concept from multivariate calculus. I'll leave it unexplained for now, but normally if we want to optimise some multivariate function $g(\vec{v})$ with the constraint that some other function $h(\vec{v})$ of the same variables is equal to a constant k, we first optimise the function $g(\vec{v}) - \lambda(h(\vec{v}))$ where λ is an arbitrary constant, and then check which solutions match our initial constraint $h(\vec{v}) = k$.[10]

It turns we need to do exactly the same thing when we're dealing with functionals instead. (Imagine the number of dimensions of the input vector $\vec{(v)}$ just tending towards infinity.)

So, we want to minimise

$$\int_0^{2\pi} \left(\frac{1}{2} f^2(\theta) - \lambda f(\theta) \right) \mathrm{d}\theta.$$

Applying the Euler-Lagrange equation (and taking care since our function here is also called f),

$$\begin{aligned} \frac{\partial}{\partial f} \left[\frac{1}{2} f^2(\theta) - \lambda f(\theta) \right] &= \frac{\mathrm{d}}{\mathrm{d}\theta} \left(\frac{\partial}{\partial f'} \left[\frac{1}{2} f^2(\theta) - \lambda f(\theta) \right] \right) \\ \implies f(\theta) - \lambda &= \frac{\mathrm{d}}{\mathrm{d}\theta}(0) \\ \implies f(\theta) &= \lambda. \end{aligned}$$

That is, the radius is some constant λ — a circle! Our constraint was that the perimeter is l and so the radius is just

$$r = \lambda = \frac{l}{2\pi}.$$

Admittedly that took a little bit of hand-waving when talking about constrained optimisation, but I hope to explain what's going on in more detail in a future document. For now, I included it because I think it's just another of many wonderful examples of things the Euler-Lagrange equation can do.

10 A specialisation: the Beltrami identity

Just before I round off, I want to mention the fact that although we derived the Euler-Lagrange equation to find extremals of functionals of the form

$$F[y] = \int_{x_1}^{x_2} f(x, y(x), y'(x))\, \mathrm{d}x,$$

so far none of our functions f within the integral have explicitly depended upon the actual parameter x at all. It turns out that in many physical problems this is true: we actually only care about the *value* of the function $y(x)$ on the interval (x_1, x_2), not the input to the function.

For such situations, where $\dfrac{\partial f}{\partial x} = 0$, we can find a simplified and less general form of the Euler-Lagrange equation.

[10] This is to do with tangency of the two functions' contour lines — it's very interesting and I hope to do a follow-up explaining this and some of the consequences in Lagrangian and Hamiltonian mechanics.

We start by looking at the total derivative of f with respect to x. The multivariate chain rule gives

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}x} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial f}{\partial y'}\frac{\mathrm{d}y'}{\mathrm{d}x} \\ &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial y'}y''\end{aligned}$$

but we're dealing with the situation when $\dfrac{\partial f}{\partial x} = 0$, so

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial y'}y''.$$

Rearranging this leads to

$$\frac{\partial f}{\partial y}y' = \frac{\mathrm{d}f}{\mathrm{d}x} - \frac{\partial f}{\partial y'}y''. \tag{16}$$

Now, the Euler-Lagrange equation says that if y minimises the value of the integral, then

$$\frac{\partial f}{\partial y} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)$$

and multiplying this by y' gives

$$\frac{\partial f}{\partial y}y' = y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right). \tag{17}$$

So, we can set the right hand sides of eqs. (16) and (17) equal! Doing so results in

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}x} - \frac{\partial f}{\partial y'}y'' &= y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \\ \implies \frac{\mathrm{d}f}{\mathrm{d}x} &= y''\frac{\partial f}{\partial y'} + y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right).\end{aligned}$$

Looking at the right hand side, though, we can apply the product rule in reverse to say

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial f}{\partial y'}\right).$$

Finally, integrating both sides with respect to x, we come to

$$\begin{aligned} f &= y'\frac{\partial f}{\partial y'} + c \\ \implies f - y'\frac{\partial f}{\partial y'} &= c\end{aligned}$$

for some constant c.

This is Beltrami's identity, and it makes finding extrema of functionals with no explicit x-dependence *much* easier than with the raw Euler-Lagrange equation: computation of only one derivative is involved and the result will generally immediately be a just first-order differential equation. In fact, we could have used this instead in all of the examples given in this paper.

11 Conclusion and questions for thought

We've come a long way and are now armed with the tools to solve all sorts of problems that would have seemed impenetrable before. The true calculus of variations in this paper was all in deriving the Euler-Lagrange equation; that's where we discussed the concepts of functionals and their variations, the concepts that are really at the heart of this field. This is the reason that the Euler-Lagrange equation is so beautiful: it empowers us to solve endless numbers of problems that should require a deep conceptual understanding of functionals, all using just a single differential equation.

I've gone through the use of the calculus of variations in finding the shortest path between two points, and in finding the path of *least time* between two points, as well as in proving the fundamental theorem of calculus and solving the Greeks' isoperimetric problem. Those are just a few of the surprisingly many areas that I've realised the field touches though, and here I include some questions for the reader to ponder. When I have some time, I might answer some of them too.

11.1 Future questions to consider

- We found the shortest path between two points on a plane is a straight line. What if the two points are on a sphere, or a pseudosphere? Investigate the effect of non-Euclidean geometry on the shortest path problem. This is useful, for example, for choosing the route that an airplane should fly between two cities on Earth.
- Even more abstractly, could we use the same technique to find the shortest path between two points on some general smooth surface? What applications could this have in real life?
- We derived the Euler-Lagrange equation for functionals of single-variable functions. Generalise the result to multivariate functions. Does one equation still apply, or do several need to be satisfied simultaneously?
- Similarly, the Euler-Lagrange equation applies to functionals that are integrals of a function of the form $f(x, y(x), y'(x))$. What if the function has dependence on higher derivatives of y? Derive a similar result for functions of the form $f(x, y, y'', y''', ...)$.
- The brachistochrone problem that we solved is accompanied by a closely related problem: is there a curve on which any frictionless particle that is released anywhere from rest will reach the bottom (the minimum) of the curve in the *same* amount of time? Such a curve is called a tautochrone. Is it also a cycloid?
- In solving the isoperimetric problem, we found the 2-dimensional shape that enclosed the largest *area*. What if we want the solid that encloses the largest volume? Solve the problem for n dimensions.
- A similar problem to the isoperimetric one is to find the volume of revolution between two points that has the minimum surface area. How is it related to the isoperimetric problem? Can this be generalised to higher dimensions as well?

- Look further into the use of Lagrange multipliers in both multivariate calculus and the calculus of variations. What is the intuition behind their effect?
- Finally, investigate the application of the calculus of variations to the principle of least action. Can you use the techniques discussed here to prove Newton's second law using just the aforementioned principle? Look into the Lagrangian and Hamiltonian reformulations of classical mechanics — they are alternatives to Newtonian mechanics based heavily on the Euler-Lagrange equation.

12 References

While I tried to derive as much of this as I could by myself (after the initial explanation of the Euler-Lagrange equation), I of course heavily referred to existing resources for inspiration and help when I got stuck. The main ones I used are listed below.

[1] Jose Figueroa-O'Farrill. *Brief Notes on the Calculus of Variations*. University of Edinburgh. URL: http://www.maths.ed.ac.uk/~jmf/Teaching/Lectures/CoV.pdf.

[2] Paul Kunkel. *The Brachistochrone*. Whistler Alley Mathematics. URL: http://whistleralley.com/brachistochrone/brachistochrone.htm.

[3] Charles Byrne. *Notes on The Calculus of Variations*. University of Massachusetts at Lowell. 2009. URL: https://pdfs.semanticscholar.org/a26c/0342b54e8456930bf4d320280c1a08f732a5.pdf.

[4] Eric W. Weisstein. *Beltrami Identity*. Wolfram MathWorld. URL: http://mathworld.wolfram.com/BeltramiIdentity.html.

[5] Markus Grasmair. *Basics of Calculus of Variations*. Norwegian University of Science and Technology. URL: https://wiki.math.ntnu.no/_media/tma4180/2015v/calcvar.pdf.

[6] Yutaka Nishiyama. *The Brachistochrone Curve: The Problem of Quickest Descent*. Osaka University of Economics. URL: http://www.osaka-ue.ac.jp/zemi/nishiyama/math2010/cycloid.pdf.