1

Calculus of variations

We begin our tour of useful mathematics with what is called the calculus of variations. Many physics problems can be formulated in the language of this calculus, and once they are there are useful tools to hand. In the text and associated exercises we will meet some of the equations whose solution will occupy us for much of our journey.

1.1 What is it good for?

The classical problems that motivated the creators of the calculus of variations include:

(i) Dido’s problem: In Virgil’s Aeneid, Queen Dido of Carthage must find the largest area that can be enclosed by a curve (a strip of bull’s hide) of fixed length.
(ii) Plateau’s problem: Find the surface of minimum area for a given set of bounding curves. A soap film on a wire frame will adopt this minimal-area configuration.
(iii) Johann Bernoulli’s brachistochrone: A bead slides down a curve with fixed ends. Assuming that the total energy $\frac{1}{2}mv^2 + V(x)$ is constant, find the curve that gives the most rapid descent.
(iv) Catenary: Find the form of a hanging heavy chain of fixed length by minimizing its potential energy.

These problems all involve finding maxima or minima, and hence equating some sort of derivative to zero. In the next section we define this derivative, and show how to compute it.

1.2 Functionals

In variational problems we are provided with an expression $J[y]$ that “eats” whole functions $y(x)$ and returns a single number. Such objects are called functionals to distinguish them from ordinary functions. An ordinary function is a map $f : \mathbb{R} \rightarrow \mathbb{R}$. A functional $J$ is a map $J : C^\infty(\mathbb{R}) \rightarrow \mathbb{R}$ where $C^\infty(\mathbb{R})$ is the space of smooth (having derivatives of all orders) functions. To find the function $y(x)$ that maximizes or minimizes a given functional $J[y]$ we need to define, and evaluate, its functional derivative.
2

1 Calculus of variations

1.2.1 The functional derivative

We restrict ourselves to expressions of the form

\[ J[y] = \int_{x_1}^{x_2} f(x, y, y', y'', \ldots, y^{(n)}) \, dx, \]  \hspace{1cm} (1.1)\]

where \( f \) depends on the value of \( y(x) \) and only finitely many of its derivatives. Such functionals are said to be local in \( x \).

Consider first a functional \( J = \int f \, dx \) in which \( f \) depends only on \( x, y \) and \( y' \). Make a change \( y(x) \rightarrow y(x) + \varepsilon \eta(x) \), where \( \varepsilon \) is a (small) \( x \)-independent constant. The resultant change in \( J \) is

\[ J[y + \varepsilon \eta] - J[y] = \int_{x_1}^{x_2} \left\{ f(x, y + \varepsilon \eta, y' + \varepsilon \eta') - f(x, y, y') \right\} \, dx \]

\[ = \int_{x_1}^{x_2} \left\{ \varepsilon \eta \frac{\partial f}{\partial y} + \varepsilon \frac{d \eta}{dx} \frac{\partial f}{\partial y'} + O(\varepsilon^2) \right\} \, dx \]

\[ = \left[ \varepsilon \eta \frac{\partial f}{\partial y} \right]_{x_1}^{x_2} + \int_{x_1}^{x_2} (\varepsilon \eta(x)) \left( \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) \right) \, dx + O(\varepsilon^2). \]

If \( \eta(x_1) = \eta(x_2) = 0 \), the variation \( \delta y(x) \equiv \varepsilon \eta(x) \) in \( y(x) \) is said to have “fixed endpoints”. For such variations the integrated-out part \( \ldots \) vanishes. Defining \( \delta J \) to be the \( O(\varepsilon) \) part of \( J[y + \varepsilon \eta] - J[y] \), we have

\[ \delta J = \int_{x_1}^{x_2} (\varepsilon \eta(x)) \left( \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) \right) \, dx. \]  \hspace{1cm} (1.2)\]

The function

\[ \frac{\delta J}{\delta y(x)} = \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) \]  \hspace{1cm} (1.3)\]

is called the functional (or Fréchet) derivative of \( J \) with respect to \( y(x) \). We can think of it as a generalization of the partial derivative \( \partial J/\partial y_i \), where the discrete subscript “\( i \)” on \( y \) is replaced by a continuous label “\( x \)”, and sums over \( i \) are replaced by integrals over \( x \):

\[ \delta J = \sum_i \frac{\partial J}{\partial y_i} \delta y_i \rightarrow \int_{x_1}^{x_2} dx \left( \frac{\delta J}{\delta y(x)} \right) \delta y(x). \]  \hspace{1cm} (1.4)\]
1.2 Functionals

1.2.2 The Euler–Lagrange equation

Suppose that we have a differentiable function $J(y_1, y_2, \ldots, y_n)$ of $n$ variables and seek its stationary points – these being the locations at which $J$ has its maxima, minima and saddle points. At a stationary point $(y_1, y_2, \ldots, y_n)$ the variation

$$
\delta J = \sum_{i=1}^{n} \frac{\partial J}{\partial y_i} \delta y_i
$$

must be zero for all possible $\delta y_i$. The necessary and sufficient condition for this is that all partial derivatives $\frac{\partial J}{\partial y_i}, i = 1, \ldots, n$ be zero. By analogy, we expect that a functional $J[y]$ will be stationary under fixed-endpoint variations $y(x) \to y(x) + \delta y(x)$, when the functional derivative $\frac{\delta J}{\delta y(x)}$ vanishes for all $x$. In other words, when

$$
\frac{\partial f}{\partial y(x)} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'(x)} \right) = 0, \quad x_1 < x < x_2.
$$

(1.6)

The condition (1.6) for $y(x)$ to be a stationary point is usually called the Euler–Lagrange equation.

That $\delta J/\delta y(x) \equiv 0$ is a sufficient condition for $\delta J$ to be zero is clear from its definition in (1.2). To see that it is a necessary condition we must appeal to the assumed smoothness of $y(x)$. Consider a function $y(x)$ at which $J[y]$ is stationary but where $\delta J/\delta y(x)$ is non-zero at some $x_0 \in [x_1, x_2]$. Because $f(y, y', x)$ is smooth, the functional derivative $\delta J/\delta y(x)$ is also a smooth function of $x$. Therefore, by continuity, it will have the same sign throughout some open interval containing $x_0$. By taking $\delta y(x) = \varepsilon \eta(x)$ to be zero outside this interval, and of one sign within it, we obtain a non-zero $\delta J$ – in contradiction to stationarity. In making this argument, we see why it was essential to integrate by parts so as to take the derivative off $\delta y$: when $y$ is fixed at the endpoints, we have $\int \delta y' \, dx = 0$, and so we cannot find a $\delta y'$ that is zero everywhere outside an interval and of one sign within it.

When the functional depends on more than one function $y$, then stationarity under all possible variations requires one equation

$$
\frac{\delta J}{\delta y_i(x)} = \frac{\partial f}{\partial y_i} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'_i} \right) = 0
$$

(1.7)

for each function $y_i(x)$.

If the function $f$ depends on higher derivatives, $y''$, $y^{(3)}$, etc., then we have to integrate by parts more times, and we end up with

$$
0 = \frac{\delta J}{\delta y(x)} = \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) + \frac{d^2}{dx^2} \left( \frac{\partial f}{\partial y''} \right) - \frac{d^3}{dx^3} \left( \frac{\partial f}{\partial y^{(3)}} \right) + \cdots.
$$

(1.8)
1 Calculus of variations

1.2.3 Some applications

Now we use our new functional derivative to address some of the classic problems mentioned in the introduction.

Example: Soap film supported by a pair of coaxial rings (Figure 1.1). This is a simple case of Plateau’s problem. The free energy of the soap film is equal to twice (once for each liquid–air interface) the surface tension $\sigma$ of the soap solution times the area of the film. The film can therefore minimize its free energy by minimizing its area, and the axial symmetry suggests that the minimal surface will be a surface of revolution about the $x$-axis. We therefore seek the profile $y(x)$ that makes the area

$$J[y] = 2\pi \int_{x_1}^{x_2} y\sqrt{1 + y'^2} \, dx$$

of the surface of revolution the least among all such surfaces bounded by the circles of radii $y(x_1) = y_1$ and $y(x_2) = y_2$. Because a minimum is a stationary point, we seek candidates for the minimizing profile $y(x)$ by setting the functional derivative $\delta J/\delta y(x)$ to zero.

We begin by forming the partial derivatives

$$\frac{\partial f}{\partial y} = 4\pi \sqrt{1 + y'^2}, \quad \frac{\partial f}{\partial y'} = \frac{4\pi yy'}{\sqrt{1 + y'^2}}$$

and use them to write down the Euler–Lagrange equation

$$\sqrt{1 + y'^2} - \frac{d}{dx} \left( \frac{yy'}{\sqrt{1 + y'^2}} \right) = 0.$$
Performing the indicated derivative with respect to \( x \) gives

\[
\sqrt{1 + y^2} - \frac{(y')^2}{\sqrt{1 + y^2}} - \frac{yy''}{\sqrt{1 + y^2}} + \frac{y(y')^2y''}{(1 + y^2)^{3/2}} = 0. \tag{1.12}
\]

After collecting terms, this simplifies to

\[
\frac{1}{\sqrt{1 + y^2}} - \frac{yy''}{(1 + y^2)^{3/2}} = 0. \tag{1.13}
\]

The differential equation (1.13) still looks a trifle intimidating. To simplify further, we multiply by \( y' \) to get

\[
0 = \frac{y'}{\sqrt{1 + y^2}} - \frac{yy'y''}{(1 + y^2)^{3/2}}
= \frac{d}{dx} \left( \frac{y}{\sqrt{1 + y^2}} \right). \tag{1.14}
\]

The solution to the minimization problem therefore reduces to solving

\[
\frac{y}{\sqrt{1 + y^2}} = \kappa, \tag{1.15}
\]

where \( \kappa \) is an as yet undetermined integration constant. Fortunately this nonlinear, first-order, differential equation is elementary. We recast it as

\[
\frac{dy}{dx} = \sqrt{\frac{y^2}{\kappa^2} - 1} \tag{1.16}
\]

and separate variables

\[
\int dx = \int \frac{dy}{\sqrt{\frac{y^2}{\kappa^2} - 1}}. \tag{1.17}
\]

We now make the natural substitution \( y = \kappa \cosh t \), whence

\[
\int dx = \kappa \int dt. \tag{1.18}
\]

Thus we find that \( x + a = \kappa t \), leading to

\[
y = \kappa \cosh \frac{x + a}{\kappa}. \tag{1.19}
\]
1 Calculus of variations

We select the constants $\kappa$ and $a$ to fit the endpoints $y(x_1) = y_1$ and $y(x_2) = y_2$.

**Example: Heavy chain over pulleys.** We cannot yet consider the form of the catenary, a hanging chain of fixed length, but we can solve a simpler problem of a heavy flexible cable draped over a pair of pulleys located at $x = \pm L$, $y = h$, and with the excess cable resting on a horizontal surface as illustrated in Figure 1.2.

The potential energy of the system is

$$P.E. = \sum mgy = \rho g \int_{-L}^{L} y \sqrt{1 + (y')^2} \, dx + \text{const.} \quad (1.20)$$

Here the constant refers to the unchanging potential energy

$$2 \times \int_{0}^{h} mg \, dy = mgh^2 \quad (1.21)$$

of the vertically hanging cable. The potential energy of the cable lying on the horizontal surface is zero because $y$ is zero there. Notice that the tension in the suspended cable is being tacitly determined by the weight of the vertical segments.

The Euler–Lagrange equations coincide with those of the soap film, so

$$y = \kappa \cosh \left( \frac{x + a}{\kappa} \right) \quad (1.22)$$

where we have to find $\kappa$ and $a$. We have

$$h = \kappa \cosh(-L + a)/\kappa,$$

$$= \kappa \cosh(L + a)/\kappa, \quad (1.23)$$
1.2 Functionals

so \( a = 0 \) and \( h = \kappa \cosh \frac{L}{\kappa} \). Setting \( t = \frac{L}{\kappa} \) this reduces to

\[
\left( \frac{h}{L} \right) t = \cosh t.
\]  
(1.24)

By considering the intersection of the line \( y = \frac{ht}{L} \) with \( y = \cosh t \) (Figure 1.3) we see that if \( h/L \) is too small there is no solution (the weight of the suspended cable is too big for the tension supplied by the dangling ends) and once \( h/L \) is large enough there will be two possible solutions. Further investigation will show that the solution with the larger value of \( \kappa \) is a point of stable equilibrium, while the solution with the smaller \( \kappa \) is unstable.

**Example:** The brachistochrone. This problem was posed as a challenge by Johann Bernoulli in 1696. He asked what shape should a wire with endpoints \((0, 0)\) and \((a, b)\) take in order that a frictionless bead will slide from rest down the wire in the shortest possible time (Figure 1.4). The problem’s name comes from Greek: βραχιστος means shortest and χρονος means time.

When presented with an ostensibly anonymous solution, Johann made his famous remark: “Tanquam ex unguem leonem” (I recognize the lion by his clawmark) meaning that he recognized that the author was Isaac Newton.

Johann gave a solution himself, but that of his brother Jacob Bernoulli was superior and Johann tried to pass it off as his. This was not atypical. Johann later misrepresented the publication date of his book on hydraulics to make it seem that he had priority in this field over his own son, Daniel Bernoulli.

We begin **our** solution of the problem by observing that the total energy

\[
E = \frac{1}{2} m(\dot{x}^2 + \dot{y}^2) - mgy = \frac{1}{2} m\dot{x}^2 (1 + \dot{y}^2) - mgy, 
\]  
(1.25)
of the bead is constant. From the initial condition we see that this constant is zero. We therefore wish to minimize

$$ T = \int_0^T dt = \int_0^a 1 \frac{dx}{\dot{x}} = \int_0^a \sqrt{\frac{1 + y'^2}{2gy}} \, dx $$ (1.26)

so as to find $y(x)$, given that $y(0) = 0$ and $y(a) = b$. The Euler–Lagrange equation is

$$ yy'' + \frac{1}{2} (1 + y'^2) = 0. $$ (1.27)

Again this looks intimidating, but we can use the same trick of multiplying through by $y'$ to get

$$ y' \left( yy'' + \frac{1}{2} (1 + y'^2) \right) = \frac{1}{2} \frac{d}{dx} \left( y(1 + y'^2) \right) = 0. $$ (1.28)

Thus

$$ 2c = y(1 + y'^2). $$ (1.29)

This differential equation has a parametric solution

$$ x = c(\theta - \sin \theta), $$

$$ y = c(1 - \cos \theta), $$ (1.30)

(as you should verify) and the solution is the cycloid shown in Figure 1.5. The parameter $c$ is determined by requiring that the curve does in fact pass through the point $(a, b)$. 

![Figure 1.4 Bead on a wire.](image-url)
1.2 Functionals

How did we know that we could simplify both the soap-film problem and the brachistochrone by multiplying the Euler equation by $y'$? The answer is that there is a general principle, closely related to energy conservation in mechanics, that tells us when and how we can make such a simplification. The $y'$ trick works when the $f$ in \( \int f \, dx \) is of the form $f(y, y')$, i.e. has no explicit dependence on $x$. In this case the last term in

$$\frac{df}{dx} = y' \frac{\partial f}{\partial y} + y'' \frac{\partial f}{\partial y'} + \frac{\partial f}{\partial x}$$  (1.31)

is absent. We then have

$$\frac{d}{dx} \left( f - y' \frac{\partial f}{\partial y'} \right) = y' \frac{\partial f}{\partial y} + y'' \frac{\partial f}{\partial y'} - y'' \frac{\partial f}{\partial y'} - y' \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right)$$

$$= y' \left( \frac{\partial f}{\partial y} - \frac{d}{dx} \left( \frac{\partial f}{\partial y'} \right) \right),$$  (1.32)

and this is zero if the Euler–Lagrange equation is satisfied.

The quantity

$$I = f - y' \frac{\partial f}{\partial y'}$$  (1.33)

is called a first integral of the Euler–Lagrange equation. In the soap-film case

$$f - y' \frac{\partial f}{\partial y'} = y\sqrt{1 + (y')^2} - \frac{y(y')^2}{\sqrt{1 + (y')^2}} = \frac{y}{\sqrt{1 + (y')^2}}.$$  (1.34)

When there are a number of dependent variables $y_i$, so that we have

$$J[y_1, y_2, \ldots, y_n] = \int f(y_1, y_2, \ldots, y_n, y_1', y_2', \ldots, y_n') \, dx$$  (1.35)
then the first integral becomes
\[ I = f - \sum_i y_i' \frac{\partial f}{\partial y_i'} . \] (1.36)

Again
\[ \frac{dl}{dx} = \frac{d}{dx} \left( f - \sum_i y_i' \frac{\partial f}{\partial y_i'} \right) \]
\[ = \sum_i \left( y_i' \frac{\partial f}{\partial y_i} + y_i'' \frac{\partial f}{\partial y_i'} - y_i' \frac{\partial f}{\partial y_i} - y_i' \frac{d}{dx} \left( \frac{\partial f}{\partial y_i'} \right) \right) \]
\[ = \sum_i y_i' \frac{\partial f}{\partial y_i} - \frac{d}{dx} \left( \frac{\partial f}{\partial y_i'} \right) , \] (1.37)
and this is zero if the Euler–Lagrange equation is satisfied for each \( y_i \).

Note that there is only one first integral, no matter how many \( y_i \)'s there are.

1.3 Lagrangian mechanics

In his *Mécanique Analytique* (1788) Joseph-Louis de La Grange, following Jean d’Alembert (1742) and Pierre de Maupertuis (1744), showed that most of classical mechanics can be recast as a variational condition: the principle of least action. The idea is to introduce the Lagrangian function \( L = T - V \) where \( T \) is the kinetic energy of the system and \( V \) the potential energy, both expressed in terms of generalized coordinates \( q^i \) and their time derivatives \( \dot{q}^i \). Then, Lagrange showed, the multitude of Newton’s \( F = ma \) equations, one for each particle in the system, can be reduced to
\[ \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}^i} \right) - \frac{\partial L}{\partial q^i} = 0, \] (1.38)
one equation for each generalized coordinate \( q^i \). Quite remarkably – given that Lagrange’s derivation contains no mention of maxima or minima – we recognize that this is precisely the condition that the action functional
\[ S[q] = \int_{t_{\text{initial}}}^{t_{\text{final}}} L(t, q^i, \dot{q}^i) \, dt \] (1.39)
be stationary with respect to variations of the trajectory \( q^i(t) \) that leave the initial and final points fixed. This fact so impressed its discoverers that they believed they had uncovered the unifying principle of the universe. Maupertuis, for one, tried to base a proof of the existence of God on it. Today the action integral, through its starring role in