MS327

Deterministic and stochastic dynamics

Unit 5

Introduction to the calculus of variations

This publication forms part of an Open University module. Details of this and other Open University modules can be obtained from Student Recruitment, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)300 303 5303; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

Note to reader
Mathematical/statistical content at the Open University is usually provided to students in printed books, with PDFs of the same online. This format ensures that mathematical notation is presented accurately and clearly. The PDF of this extract thus shows the content exactly as it would be seen by an Open University student. Please note that the PDF may contain references to other parts of the module and/or to software or audio-visual components of the module. Regrettably mathematical and statistical content in PDF files is unlikely to be accessible using a screenreader, and some OpenLearn units may have PDF files that are not searchable. You may need additional help to read these documents.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2016.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Latimer Trend and Company Ltd, Plymouth.

ISBN 978 1 8487 3962 8

1.1

Contents

Introduction

In Book 1, we introduced various types of dynamical systems. These systems could be classified as:

- *conservative*, where the total energy of the system is constant (i.e. conserved) throughout the motion
- *non-conservative*, where energy is not conserved, as for example in the presence of forcing or damping
- *deterministic*, where the equations describing the motion have no randomness in them
- *non-deterministic* (or *random* or *stochastic*), where the equations describing the motion have randomness built in, as for example in a random walk.

In Book 2, we explore some aspects of conservative and non-conservative *deterministic* systems in more detail, before moving on to discuss random systems in Book 3.

In earlier modules, you may have seen how to derive the equations of motion of simple mechanical systems using Newton's laws of motion. In Unit 6, we will introduce you to another method called *Lagrangian mechanics*, which applies to conservative systems, and is arguably much easier to use and more powerful than Newtonian mechanics – although perhaps less intuitive since it is based on energies rather than forces. However, before we get to Unit 6, we first need to set up some important mathematical apparatus called the *calculus of variations*: this is the main purpose of this unit.

In ordinary calculus, we often work with real functions, which are rules for mapping real numbers to real numbers; for example, the function $\sin x$ maps the whole of the real line to the interval $[-1, 1]$. Functions can have various properties: for instance, they can be continuous and differentiable, and they can have stationary points and local maxima and minima.

In the calculus of variations, we work with *functionals*: these are objects that map functions to the real line. For example,

$$S[f] = \int_0^1 (f'(x))^2 \, dx$$

is a functional, which for any sufficiently well-behaved real function f gives us a real number $S[f]$. (Here and throughout this unit and the next, $f'(x)$ denotes the derivative df/dx evaluated at x.) The square bracket notation $S[f]$ is used to emphasise the fact that the functional S depends on the choice of function.

Functionals share many of the same properties as functions. In particular, the notion of a stationary point of a function has an important analogue in the theory of functionals, which gives rise to the calculus of variations, as you will see. The calculus of variations is a hugely important topic in the natural sciences. It leads naturally to the Lagrangian formulation of mechanics, mentioned above, from which most dynamical equations of mathematical physics can be derived. It is also a powerful mathematical tool, finding applications in subjects as diverse as statics, optics, differential geometry, approximate solutions of differential equations and control theory. In addition to this, the calculus of variations is an active topic of study in its own right.

The structure of this unit is as follows. In Section 1 we introduce many of the key ingredients of the calculus of variations, by solving a seemingly simple problem – finding the shortest distance between two points in a plane. In particular, this section introduces the notion of a functional and that of a *stationary path*. In Section 2 we briefly describe a few basic problems that can be formulated in terms of functionals, in order to give you a feel for the range of problems that can be solved using the calculus of variations; *this section will not be assessed*. Section 3 is a short interlude about partial and total derivatives, which are used extensively throughout the rest of the unit. Section 4 is the most important section of this unit. It contains a derivation of the *Euler–Lagrange equation*, which will be used throughout the rest of this unit and the next. Finally, in Section 5 we apply the Euler–Lagrange equation to solve some of the problems discussed in Section 2, as well as a problem arising from a new topic, called *Fermat's principle*, which serves as a bridge to the next unit.

1 Shortest distance between two points

In this section we take a swift tour through variational principles by considering perhaps the simplest physical example possible – the shortest distance between two points in a plane.

You may think that this is a somewhat trivial example, since the shortest path between two points in a plane is the straight line joining them. However, it is almost always easiest to understand a new idea by applying it to a simple familiar problem, so here we introduce the ideas of the calculus of variations by proving this trivial fact. The algebra involved may seem over-complicated for such a basic problem, but far more complicated problems (like the problem of finding the shortest distance between two points on a curved surface) can be solved using the same principles.

We begin in Subsection 1.1 by showing that the distance between two points can be expressed as a functional. Then in Subsection 1.2 we show that the shortest distance between two points in a plane is a straight line.

1.1 Distance between two points on a given curve

First, we need an expression for the length of a curve between two given points in a plane.

Suppose that we are given two points P_a and P_b with Cartesian coordinates (a, A) and (b, B), respectively. Furthermore, suppose that $y = y(x)$, where $a \leq x \leq b$, is a smooth curve that joins P_a at $x = a$ to P_b at $x = b$, so that $y(a) = A$ and $y(b) = B$, as shown in Figure 1.

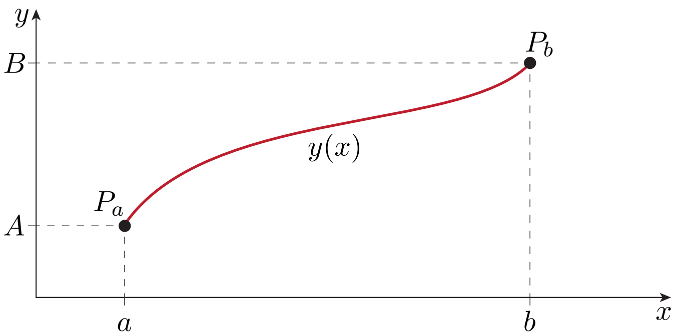


Figure 1 Graph of the curve $y = y(x)$ passing through points P_a and P_b

You may have derived a formula for the length of this curve in earlier modules. It is obtained by dividing the interval $a \leq x \leq b$ into N intervals of length $\delta x = (b - a)/N$. Let us denote the endpoints of these intervals by $x_0, x_1, \ldots, x_N$, in that order, so that $x_0 = a$, $x_N = b$ and $x_{k+1} - x_k = \delta x$ for $k = 0, 1, \ldots, N - 1$, as shown in Figure 2.

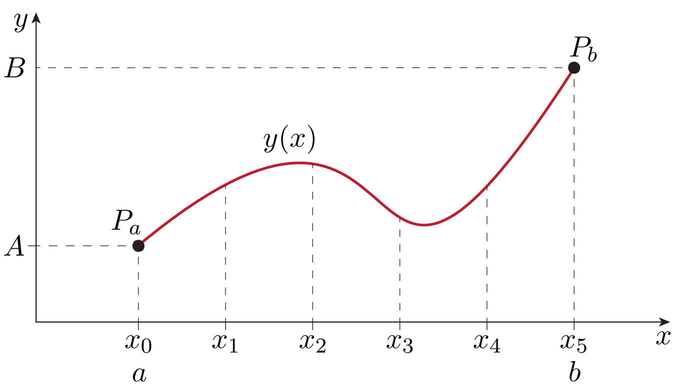


Figure 2 Subdivision of the interval $[a, b]$ into N equal intervals; in this case $N = 5$

The curve is now approximated by a sequence of N straight-line segments. Let δs_k be the length of the segment above the interval $[x_k, x_{k+1}]$. This segment is illustrated in Figure 3.

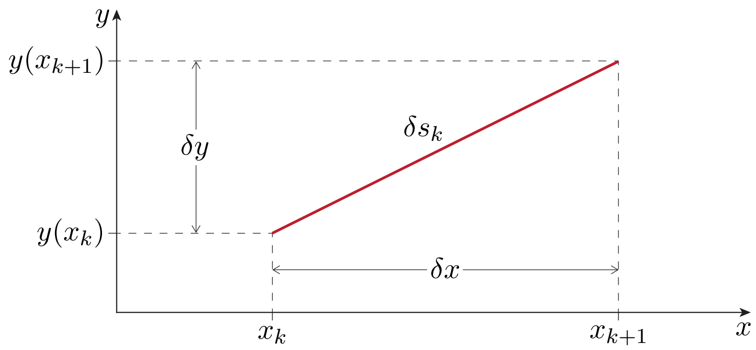


Figure 3 Segment of the curve $y = y(x)$ between x_k and x_{k+1}

The length of each segment can be determined using Pythagoras' theorem:

$$\delta s_k = \sqrt{\delta x^2 + \delta y^2} = \delta x \sqrt{1 + \left(\frac{\delta y}{\delta x}\right)^2}.$$

However, $\delta y/\delta x$ is just the gradient of the segment, which in the limit as $N \to \infty$ is given by $\delta y/\delta x = y'(x_k)$ (i.e. the derivative dy/dx of y evaluated at $x = x_k$). Therefore

$$\delta s_k = \delta x \sqrt{1 + y'(x_k)^2}. \tag{1}$$

The approximate distance from (a, A) to (b, B) along the curve $y = y(x)$ is given by the sum of all the segment lengths:

$$S[y] \simeq \sum_{k=0}^{N-1} \delta s_k = \sum_{k=0}^{N-1} \delta x \sqrt{1 + y'(x_k)^2}.$$

In the limit as $N \to \infty$, this sum becomes the integral

$$S[y] = \int_a^b \sqrt{1 + y'(x)^2}\, dx.$$

The above expression gives the length of the curve $y = y(x)$ between the points P_a and P_b. The length changes for different functions $y(x)$. For instance, if $P_a = (-1, 0)$ and $P_b = (1, 0)$, then the distance from P_a to P_b along the straight line joining these points is 2, whereas the distance is π along the semicircle of which this straight line is a diameter.

We denote the numerical value of the length of the curve by $S[y]$, which emphasises the fact that the length depends on the function $y(x)$. We do not write $S(y)$, because we wish to distinguish $S[y]$ from the real-valued function $y(x)$. The quantity $S[y]$ is our first example of a *functional*; it maps functions $y(x)$ that satisfy $y(a) = A$ and $y(b) = B$ to the length of the curve $y = y(x)$ between $x = a$ and $x = b$. More generally, a functional is *any* map from functions to real numbers. Sometimes the set of functions on which a functional acts is restricted, as in the example that we have just considered.

A **functional** $S[f]$ is a map from functions f to the real numbers.

Distance along a curve

The distance between two points (a, A) and (b, B) along a curve $y = y(x)$ that passes between them is given by the functional

$$S[y] = \int_a^b \sqrt{1 + y'(x)^2}\, dx, \quad y(a) = A,\ y(b) = B. \tag{2}$$

Exercise 1

Let $P_a = (-1, 0)$ and $P_b = (1, 0)$. Verify the following statements.

(a) If $y = y(x)$ is the straight line joining P_a to P_b (i.e. the red line in Figure 4), then the functional (2) has the value $S[y] = 2$.

(b) If $y = y(x)$ is the upper semicircle with diameter the straight line from P_a to P_b (i.e. the blue curve in Figure 4), so that $x^2 + y^2 = 1$, then $S[y] = \pi$.

(*Hint*: You may assume the following standard integral, given in the Handbook: $\displaystyle\int \frac{1}{\sqrt{1 - x^2}}\, dx = \arcsin x$.)

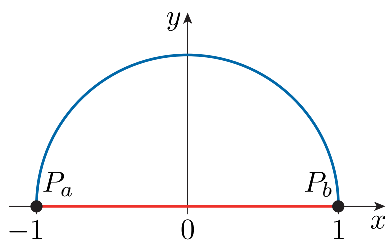


Figure 4 The two paths in Exercise 1

Exercise 2

Determine the value of the functional

$$S[y] = \int_0^1 y'(x)^2\, dx$$

for the following functions.

(a) $y(x) = x$

(b) $y(x) = x^3$

1.2 Stationary paths

In this subsection we consider how the length of the path varies with $y(x)$, and we define the notion of a *stationary path*; we also show that the straight line through the endpoints is a stationary path.

Recall that in ordinary calculus, a stationary point x of a function $y(x)$ is a point at which $y'(x) = 0$, that is, a point x for which the tangent to the graph of $y(x)$ at x is parallel to the x-axis. Once a stationary point has been found, further work is required to determine whether the function has a maximum, a minimum or a point of inflection there.

We want to define something analogous to this for functionals. To do this, let us think of a concrete problem – finding the path with *shortest* distance between the points (a, A) and (b, B) in the plane. Our problem then is to find $y = y(x)$ that will make

$$S[y] = \int_a^b \sqrt{1 + y'(x)^2}\, dx, \quad y(a) = A,\ y(b) = B,$$

as small as possible.

Suppose that $y = y(x)$ is the desired 'minimal path'. Then let $g(x)$ be any other (smooth) function that satisfies

$$g(a) = g(b) = 0, \quad \text{but otherwise } g(x) \text{ is an arbitrary function of } x.$$

Then we can define a new path by

$$\widetilde{y}(x, \varepsilon) = y(x) + \varepsilon\, g(x), \quad \text{where } \varepsilon \text{ is a real number.}$$

For each value of ε, this defines a new path $\widetilde{y}$ that also passes through the points (a, A) and (b, B), because $\widetilde{y}(a, \varepsilon) = y(a) + \varepsilon\, g(a) = A$ and $\widetilde{y}(b, \varepsilon) = y(b) + \varepsilon\, g(b) = B$. As g was defined arbitrarily (apart from its endpoints), every sufficiently well-behaved curve from a to b can be written in this way. The situation is illustrated in Figure 5.

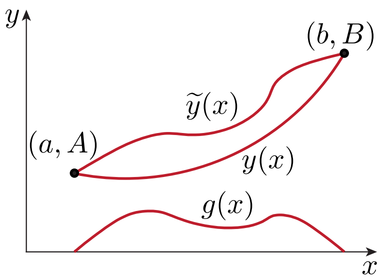


Figure 5 A minimal length path $y(x)$ and a new path $\widetilde{y}(x) = y(x) + \varepsilon\, g(x)$, both passing through the points (a, A) and (b, B)

Now, the length of the new path $\widetilde{y}$ is given by

$$\begin{aligned} S[y + \varepsilon g] &= \int_a^b \sqrt{1 + ((y + \varepsilon g)')^2}\, dx \\ &= \int_a^b \sqrt{1 + (y' + \varepsilon g')^2}\, dx, \end{aligned} \tag{3}$$

and for fixed y and g, the functional $S[y + \varepsilon g]$ gives a different path length for each value of ε. That is, for fixed y and g, $S[y + \varepsilon g]$ takes as input a real number ε and outputs the length of $y + \varepsilon g$, another real number, so $S[y + \varepsilon g]$ is a real-valued function. Furthermore, since y is our desired minimal path, $S[y + \varepsilon g]$ must take its minimum value at $\varepsilon = 0$. Mathematically, this translates to

$$\left.\frac{d}{d\varepsilon} S[y + \varepsilon g]\right|_{\varepsilon=0} = 0,$$

and this must be true for *all* functions $g(x)$ that satisfy $g(a) = g(b) = 0$. (The notation $f(x)|_{x=a}$ means evaluate $f(x)$ at $x = a$; that is, evaluate $f(a)$.) It should be clear that this condition guarantees only that $S[y + \varepsilon g]$ is stationary at $\varepsilon = 0$, and further work is required to determine whether this is a maximum, a minimum or a point of inflection. A function $y(x)$ that satisfies this equation is said to be a *stationary path* of S.

Although we have derived this result for functionals that are path lengths, the definition of stationary paths generalises to all functionals that we consider in this module.

Stationary path of a functional

Let $S[y]$ be a functional that maps functions y that satisfy $y(a) = A$ and $y(b) = B$ to the real numbers. Any such function $y(x)$ for which

$$\frac{d}{d\varepsilon} S[y + \varepsilon g]\bigg|_{\varepsilon=0} = 0, \tag{4}$$

for all functions $g(x)$ that satisfy $g(a) = g(b) = 0$, is said to be a **stationary path** of S, or alternatively a **stationary curve** or a **stationary function** of S.

Also, if S and $y(x)$ satisfy (4), then we say that S is **stationary** at $y(x)$, and sometimes we abbreviate this by merely saying that $S[y]$ is stationary.

Let us now use equation (4) to find the path between two points in a plane with shortest length. From equation (3) we have

$$\frac{d}{d\varepsilon} S[y + \varepsilon g]\bigg|_{\varepsilon=0} = \left(\frac{d}{d\varepsilon}\int_a^b \sqrt{1 + (y' + \varepsilon g')^2}\, dx\right)\bigg|_{\varepsilon=0}.$$

The integration limits a and b are independent of ε, and the order of integration and differentiation can be interchanged using Leibniz's integral rule, which we state (slightly informally) here.

Leibniz's integral rule

For any sufficiently well-behaved function $f(x, \varepsilon)$, we have

$$\frac{d}{d\varepsilon}\left(\int_a^b f(x, \varepsilon)\, dx\right) = \int_a^b \frac{d}{d\varepsilon} f(x, \varepsilon)\, dx,$$

where a and b are fixed constants (that do not depend on ε).

Applying this rule gives

$$\begin{aligned}\frac{d}{d\varepsilon} S[y + \varepsilon g]\bigg|_{\varepsilon=0} &= \int_a^b \left(\frac{d}{d\varepsilon}\sqrt{1 + (y' + \varepsilon g')^2}\right)\bigg|_{\varepsilon=0} dx \\ &= \int_a^b \left(\frac{(y' + \varepsilon g')g'}{\sqrt{1 + (y' + \varepsilon g')^2}}\right)\bigg|_{\varepsilon=0} dx \\ &= \int_a^b \frac{y'g'}{\sqrt{1 + (y')^2}}\, dx.\end{aligned}$$

If $y(x)$ is a stationary path of S, then it follows, by definition, that

$$\int_a^b \frac{y'(x)}{\sqrt{1 + y'(x)^2}}\, g'(x)\, dx = 0 \tag{5}$$

for *all* functions $g(x)$ for which $g(a) = g(b) = 0$.

To solve this equation, we integrate by parts:

$$\int_a^b u(x)\,g'(x)\,dx = \big[u(x)\,g(x)\big]_a^b - \int_a^b u'(x)\,g(x)\,dx,$$

with $u = y'(x)/\sqrt{1+y'(x)^2}$. The first term on the right-hand side vanishes because $g(a) = g(b) = 0$, so we get

$$\int_a^b ug'\,dx = -\int_a^b u'g\,dx,$$

and hence equation (5) becomes

$$\int_a^b \left(\frac{y'(x)}{\sqrt{1+y'(x)^2}}\right)' g(x)\,dx = 0.$$

Because g is arbitrary (apart from the restriction $g(a) = g(b) = 0$), this integral can vanish only if

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

(In Section 4 we will give a more careful explanation of why this is so.) Integrating both sides with respect to x gives

$$\frac{y'(x)}{\sqrt{1+y'(x)^2}} = \alpha, \tag{6}$$

for some constant α. We could do some algebra here to solve for y' in terms of α, but this is not necessary as the left-hand side of equation (6) can be constant only if

$$y'(x) = m,$$

where m is some function of α – another constant.

Integration now gives the general solution

$$y(x) = mx + c,$$

for yet another constant c; this is the equation of a straight line, as expected. The constants m and c are determined by the condition that the straight line passes through P_a and P_b; that is, $y(a) = A$ and $y(b) = B$. It is straightforward to show that this gives $m = (B - A)/(b - a)$ and $c = (Ab - Ba)/(b - a)$, so

$$y(x) = \frac{B-A}{b-a}x + \frac{Ab-Ba}{b-a}. \tag{7}$$

This analysis shows that the functional S defined in equation (2) is stationary along the straight line joining P_a to P_b.

Let us summarise what we have just learned.

Stationary path of the curve-length functional

The functional

$$S[y] = \int_a^b \sqrt{1 + y'(x)^2}\, dx, \quad y(a) = A,\ y(b) = B. \tag{8}$$

gives the length of the curve $y = y(x)$ that passes between the points $P_a = (a, A)$ and $P_b = (b, B)$. The stationary path y of this functional satisfies

$$y'(x) = \text{constant}.$$

Solving this differential equation and applying the boundary conditions $y(a) = A$ and $y(b) = B$ gives the equation of a straight line through P_a and P_b, which is the stationary path.

The main features of the procedure that you have just seen for finding the stationary path between two points are common to all variational problems in this module. That is, we start with some functional and some boundary conditions on the functions that it maps; we use equation (4) to find the functions $y(x)$ that make it stationary; this leads to a differential equation for the stationary function; then we solve the differential equation subject to the boundary conditions to find the stationary function.

Classification of stationary paths

We have not shown that a solution $y(x)$ of equation (4) is a path of *minimum* distance between P_a and P_b. To do this, we must proceed in the usual manner of ordinary calculus; that is, we must show that

$$\frac{d^2}{d\varepsilon^2} S[y + \varepsilon g]\bigg|_{\varepsilon=0} > 0.$$

We omit the algebra involved as it is tedious, and in any case, it is physically obvious that a straight line gives the shortest distance.

More generally, our ultimate goal is in applying variational principles to dynamics where one is interested only in finding functions that make the functional stationary, not in finding maxima or minima. So we concentrate on the former task.

In the next example and exercise we find the stationary paths of two more functionals, in the same way that we found the stationary path of the path length functional. In practice we rarely use this method to determine stationary paths. Instead, we use the powerful Euler–Lagrange equation, which is discussed in Section 4. To appreciate the derivation of the Euler–Lagrange equation, however, you first need to understand the method used in this section. This method will not be assessed in TMAs or the examination.

Example 1

Let $B \geq -1$. Consider the functional

$$S[y] = \int_0^1 \sqrt{1 + y'(x)}\, dx, \quad y(0) = 0,\ y(1) = B.$$

(a) Show that on a stationary path $y(x)$ of S, we have $y'(x) = \text{constant}$.

(b) Hence show that the stationary path is the straight line $y(x) = Bx$.

(c) Also show that the value of the functional on this line is $S[y] = \sqrt{1+B}$.

Solution

(a) In this case, the endpoints of the paths are (a, A) and (b, B), where $a = 0$, $A = 0$ and $b = 1$. In order to find the stationary function, we need to compute equation (4):

$$\left.\frac{d}{d\varepsilon} S[y + \varepsilon g]\right|_{\varepsilon=0} = 0, \quad \text{where } g(0) = g(1) = 0.$$

Here

$$\begin{aligned} S[y + \varepsilon g] &= \int_0^1 \sqrt{1 + (y + \varepsilon g)'}\, dx \\ &= \int_0^1 \sqrt{1 + y' + \varepsilon g'}\, dx, \end{aligned}$$

so

$$\left.\frac{d}{d\varepsilon} S[y + \varepsilon g]\right|_{\varepsilon=0} = \left.\left(\frac{d}{d\varepsilon} \int_0^1 \sqrt{1 + y' + \varepsilon g'}\, dx\right)\right|_{\varepsilon=0}.$$

The integration limits 0 and 1 are independent of ε, so the order of integration and differentiation can be interchanged using Leibniz's integral rule. This gives

$$\begin{aligned} \left.\frac{d}{d\varepsilon} S[y + \varepsilon g]\right|_{\varepsilon=0} &= \int_0^1 \left.\left(\frac{d}{d\varepsilon}\sqrt{1 + y' + \varepsilon g'}\right)\right|_{\varepsilon=0} dx \\ &= \frac{1}{2}\int_0^1 \left.\left(\frac{g'}{\sqrt{1 + y' + \varepsilon g'}}\right)\right|_{\varepsilon=0} dx \\ &= \frac{1}{2}\int_0^1 \frac{g'}{\sqrt{1 + y'}}\, dx. \end{aligned}$$

If $S[y]$ is stationary, then it follows, by definition, that

$$\int_0^1 \frac{g'(x)}{\sqrt{1 + y'(x)}}\, dx = 0$$

for all functions $g(x)$ for which $g(0) = g(1) = 0$. To solve this equation, we integrate by parts:

$$\int_0^1 u(x)\, g'(x)\, dx = \big[u(x)\, g(x)\big]_0^1 - \int_0^1 u'(x)\, g(x)\, dx,$$

with $u = 1/\sqrt{1 + y'}$.

However, the first term on the right-hand side vanishes because $g(0) = g(1) = 0$, so we get

$$\int_0^1 \left(\frac{1}{\sqrt{1 + y'(x)}} \right)' g(x)\, dx = 0.$$

Because of the freedom in choosing g, this integral can vanish only if $(1/\sqrt{1 + y'(x)})' = 0$. Integrating both sides with respect to x gives

$$\frac{1}{\sqrt{1 + y'(x)}} = \alpha,$$

where α is a constant. Rearranging this equation, we obtain $y' = m$, for some constant m.

(b) Integrating the equation $y' = m$ gives $y = mx + c$, for some constant c. The boundary conditions $y(0) = 0$ and $y(1) = B$ then give $y = Bx$, which is the equation of the stationary path.

(c) With this value for $y(x)$, the functional is

$$S[y] = \int_0^1 \sqrt{1 + y'(x)}\, dx = \int_0^1 \sqrt{1 + B}\, dx = \sqrt{1 + B}.$$

This is real if $B \geq -1$.

Exercise 3

Consider the functional

$$S[y] = \int_1^2 x(y')^2\, dx, \quad y(1) = 0, \ y(2) = 1.$$

(a) Show that the stationary path satisfies $x\, y'(x) = \text{constant}$.

(b) Hence show that the stationary function is $y(x) = \ln x / \ln 2$.

---

2 Some examples of functionals

In this section we briefly describe a few problems that can be formulated in terms of functionals (some of which are derived later), and which have solutions that are stationary paths of the functionals. The list of problems illustrates some of the types of question for which variational principles are useful, and exposes you to the sorts of functionals that are commonly used. The examples have been chosen because they require no specialist scientific knowledge to understand them. As such, the list is by no means exhaustive, and you should be aware that variational principles are applied to a much wider class of problems than those considered here.

There are no exercises for this section as *it will not be assessed*, but it is important for you to read it in order to understand the sections that follow.

2.1 Brachistochrone

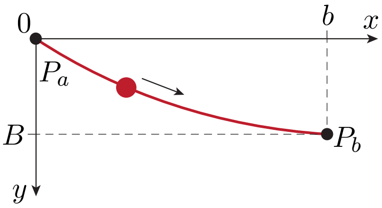


Figure 6 A bead sliding from point P_a to point P_b

The problem here is to find the shape of a wire joining two given points, so that a bead will slide down the wire under gravity from one point to the other (without friction) in the shortest time (see Figure 6).

The name given to this curve is the **brachistochrone**, which comes from the Greek *brachystos*, meaning shortest, and *chronos*, meaning time. If the y-axis is vertical, and the two given points at the ends of the wire are $P_a = (0,0)$ and $P_b = (b,B)$, and the particle starts from rest, then it can be shown that the time taken along the curve $y(x)$ is

$$T[y] = \frac{1}{\sqrt{2g}} \int_0^b \sqrt{\frac{1+(y')^2}{y}}\, dx, \quad y(0)=0,\ y(b)=B, \tag{9}$$

where g is the magnitude of the acceleration due to gravity. The map T is a functional of the wire shape $y(x)$: it maps functions $y(x)$ to the times taken for beads to descend them. The problem then is to find the function $y(x)$ that minimises T.

In Subsection 5.2 we will derive the functional T and obtain the equations for the brachistochrone, the curve that minimises T. The curve turns out to be a segment of a **cycloid**, which is the curve traced out by a point on the rim of a circular wheel rolling in a straight line (see Figure 8).

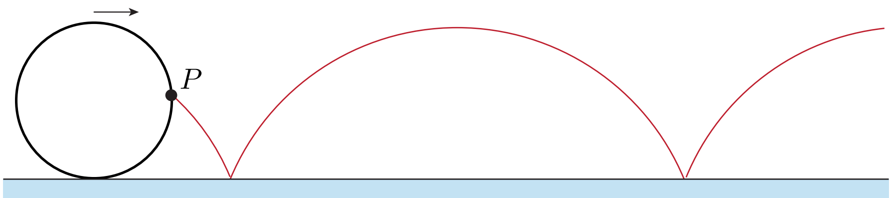


Figure 7 A cycloid: the path of a point P on the rim of a wheel

History of the brachistochrone

The problem discussed in this subsection was first considered by Galileo Galilei (1564–1642) in 1638, but, lacking the necessary mathematical techniques, he concluded erroneously that the solution is the arc of a circle passing vertically through P_a.

It was Johann Bernoulli (1667–1748) who made the problem famous when in June 1696 he challenged readers of the scientific journal *Acta Eruditorum* to solve it, reassuring them that the curve was well known to geometers. He also stated that he would demonstrate the solution at the end of the year, provided that no one else had.

In December 1696, Bernoulli extended the time limit to Easter 1697, though by this time he was in possession of the solution by Gottfried Wilhelm Leibniz (1646–1716), sent in a letter dated 16 June 1696 – Leibniz having received notification of the problem on 9 June. Isaac Newton (1642–1727) also solved the problem quickly, apparently on the day of receipt, and published his solution anonymously.

2.2 Minimal surfaces of revolution

Here the problem is to find the surface of minimal area that is generated by revolving a curve $y(x)$ about the x-axis, where $y(x)$ passes through two given points $P_a = (a, A)$ and $P_b = (b, B)$, as shown in Figure 8.

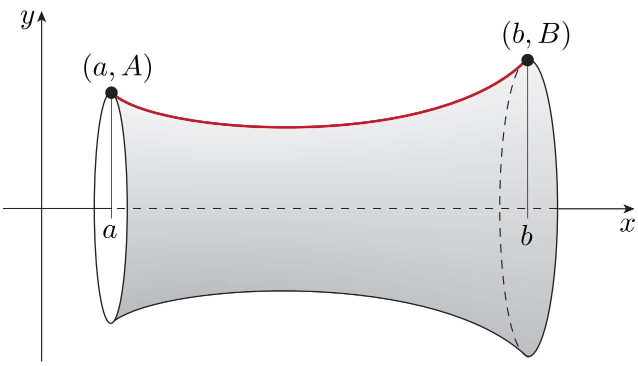


Figure 8 Cylindrical-type shape produced when a curve $y(x)$ is rotated about the x-axis

The area of this surface is shown in Subsection 5.3 to be

$$S[y] = 2\pi \int_a^b y\sqrt{1 + (y')^2}\, dx, \quad y(a) = A,\ y(b) = B. \tag{10}$$

The map $S[y]$ is a functional: it maps curves $y(x)$ between P_a and P_b to the areas of the surfaces formed by rotating such curves about the x-axis. The problem is to find the curve that minimises S.

We will see that this problem has solutions that can be expressed in terms of differentiable functions only for certain combinations of A, B and $b - a$.

Soap films

Many of us will have created soap film bubbles by dipping a loop of wire into soap solution and blowing on it. It transpires that there is a close connection between soap film surfaces and problems in the calculus of variations.

To a good approximation, a soap film will try to minimise its surface area, consistent with any wire boundary that contains it; it does this because it wants to minimise its energy, which is proportional to its surface area.

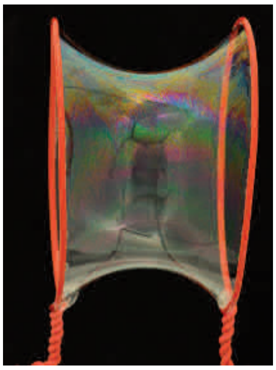


Figure 9 Minimal surface of revolution formed by a soap film

For example, a soap film supported by a circular loop of wire will be a flat circular disc.

The minimal surface of revolution can be created by dipping a pair of wire hoops into a soap solution, and gently pulling them apart. The soap film will be suspended between the two hoops, and the surface that it forms will be the 'minimal surface of revolution', as shown in Figure 9.

2.3 A problem in navigation

Consider a river with straight, parallel banks a distance a apart, and a boat that can travel with constant speed c in still water. The problem is to cross the river in the shortest time, starting and landing at given points, when there is a current.

We choose the y-axis to be the left bank, the line $x = a$ to be the right bank, and the starting point to be the origin. The water is assumed to be moving parallel to the banks with speed $v(x)$, a known function of the distance from the left bank. Then the time of passage along the path $y(x)$ from the point $(0,0)$ on the left bank to the point (a, A) on the right bank, assuming that $c > \max(v(x))$, can be shown to be

$$T[y] = \int_0^a \frac{\sqrt{c^2(1+(y')^2) - v(x)^2} - v(x)\,y'}{c^2 - v(x)^2}\,dx, \quad y(0) = 0,\ y(a) = A. \tag{11}$$

The map T is a functional: it maps paths $y(x)$ to the times to cross the river, and the problem is to find paths that minimise $T[y]$. Notice that T also depends explicitly on the speed profile $v(x)$. This problem is illustrated in Figure 10.

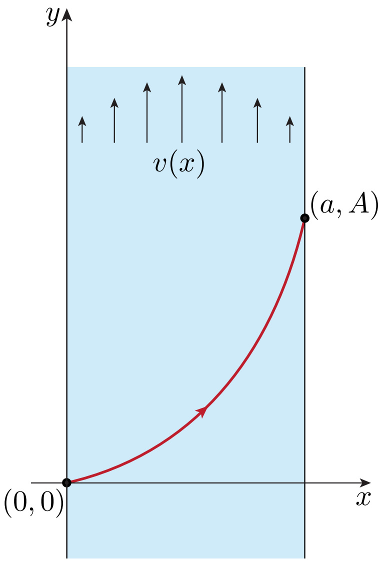


Figure 10 Path of a boat across a flowing river

2.4 Catenary

A **catenary** is the shape assumed by an inextensible chain of uniform density hanging between supports at both ends. In Figure 11 we show an example of such a curve when the points of support, $(-a, A)$ and (a, A), have the same height.

If the shape of the chain is described by the function $y(x)$, then it can be shown that the potential energy E of the chain is proportional to the functional

$$S[y] = \int_{-a}^{a} y\sqrt{1+(y')^2}\,dx, \quad y(-a) = A,\ y(a) = A, \tag{12}$$

which you encountered earlier, in equation (10).

The function y that minimises this functional, subject to the length of the curve

$$L = \int_{-a}^{a} \sqrt{1+(y')^2}\,dx$$

remaining constant, is the shape assumed by the hanging chain.

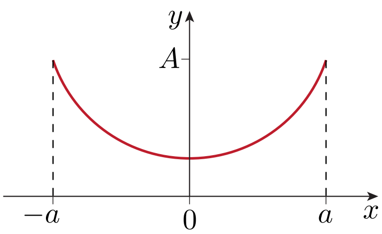


Figure 11 The catenary formed by a uniform chain hanging between two points at the same height

It turns out that the shape of the catenary is given by $y = c\cosh(x/c) + d$, where c and d are constants determined by the length of the chain and the values of a and A. Proving this, however, requires minimisation of functionals subject to constraints, which is beyond the scope of this module.

3 Reminder about partial and total derivatives

In the next section we will extend the kind of analysis that we performed in Subsection 1.2 to more general types of functionals. To do this, we must consider functions of an independent variable x, and of dependent variables $y(x)$ and $y'(x) = dy/dx$, for example $F(x, y, y') = xy^2 + 3y'$, and we will have to differentiate such functions with respect to all three variables. Although differentiating such functions is straightforward, there are aspects of the process that can cause confusion at first. In this short section we explain what is required.

Consider a function $F(x, u, v)$ of three variables; for instance,

$$F(x, u, v) = xu^2 + 3v.$$

Remember that each of the *partial derivatives* of a function of several variables is obtained by differentiating with respect to one variable while holding the others constant. The partial derivatives of $F(x, u, v)$ are

$$\frac{\partial F}{\partial x} = u^2, \quad \frac{\partial F}{\partial u} = 2xu \quad \text{and} \quad \frac{\partial F}{\partial v} = 3.$$

If we make the simple substitutions $u \to y(x)$ and $v \to y'(x)$, and treat y and y' as independent variables, then we obtain the function

$$F(x, y, y') = xy^2 + 3y',$$

and the partial derivatives

$$\frac{\partial F}{\partial x} = y^2, \quad \frac{\partial F}{\partial y} = 2xy \quad \text{and} \quad \frac{\partial F}{\partial y'} = 3.$$

Next, remember that the *total derivative* dF/dx is the rate of change of F with respect to x, without holding the other variables constant. Because y and y' depend on x, we must use the chain rule to find the total derivative of $F(x, y, y')$:

$$\begin{aligned}\frac{d}{dx}F(x, y, y') &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} + \frac{\partial F}{\partial y'}\frac{dy'}{dx} \\ &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y''.\end{aligned}$$

So in this case we obtain

$$\frac{dF}{dx} = y^2 + 2xyy' + 3y''. \tag{13}$$

We can check that this expression gives the right result for particular choices of $y(x)$. For example, suppose that $y = x^3$; then it follows that $y' = 3x^2$ and $y'' = 6x$. Substituting these values into the formula $F(x, y, y') = xy^2 + 3y'$ gives

$$F(x, x^3, 3x^2) = x^7 + 9x^2, \quad \text{so} \quad \frac{dF}{dx} = 7x^6 + 18x.$$

On the other hand, equation (13) gives

$$\frac{dF}{dx} = (x^3)^2 + 2x(x^3)(3x^2) + 18x = 7x^6 + 18x,$$

so the two methods for finding the total derivative agree.

Here are a few exercises for you to try.

Exercise 4

Let $F = x\sqrt{y^2 + (y')^2}$, where y and y' are functions of x.

(a) Find $\dfrac{\partial F}{\partial x}$, $\dfrac{\partial F}{\partial y}$ and $\dfrac{\partial F}{\partial y'}$, and hence find $\dfrac{dF}{dx}$.

(b) If $y = \sin x$, then find $\dfrac{dF}{dx}$ by

(i) substituting $y = \sin x$ into the formula for F and differentiating

(ii) using the result from part (a).

Exercise 5

For the function $F = \sqrt{x^2 + y(y')^2}$, where y and y' are functions of x, find

$$\frac{\partial F}{\partial x}, \quad \frac{\partial F}{\partial y}, \quad \frac{\partial F}{\partial y'} \quad \text{and} \quad \frac{dF}{dx}.$$

Exercise 6

For the function $F = x^2y^3$, where y is a function of x, find

$$\frac{\partial F}{\partial x}, \quad \frac{\partial F}{\partial y} \quad \text{and} \quad \frac{dF}{dx}.$$

Also, show that

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{dF}{dx}\right).$$

The formula

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{dF}{dx}\right)$$

established in Exercise 6 is valid for any sufficiently well-behaved function $F(x, y)$, where y is a function of x.

4 Euler–Lagrange equation

In this section we extend the analysis of Subsection 1.2 to find the stationary paths of more general functionals. In particular, we derive the *Euler–Lagrange equation*, which is the main result of this unit. The importance of the Euler–Lagrange equation is that it can be used to find stationary paths of a wide class of functionals without going through the type of analysis that we performed in Subsection 1.2.

4.1 Derivation of the Euler–Lagrange equation

We have seen that the functionals given in equations (9)–(12) describe a variety of physical situations. Furthermore, all of these functionals can be written in a unified way:

$$S[y] = \int_a^b F(x, y, y')\,dx, \quad y(a) = A,\ y(b) = B, \tag{14}$$

where $F(x, y, y')$ is an expression containing *at least one* of the functions y and y', and possibly also the variable x. (In contrast, notice that only y appears as an argument of the functional S; that is, we write $S[y]$.)

For example, equation (10) for the area of a surface of revolution has

$$F = 2\pi y\sqrt{1 + (y')^2},$$

so F is a function of y and y', but not of x. Also, equation (11) for the time of navigation has

$$F = \frac{\sqrt{c^2(1 + (y')^2) - v(x)^2} - v(x)\,y'}{c^2 - v(x)^2},$$

so F is a function of y and y', and also of x because of the additional function $v(x)$, which measures water speed (for example, we may know that $v(x) = x(k - x)$, for some constant k).

With all these problems, we need to find a path $y(x)$ that makes the functional (14) stationary. To do this, we proceed in exactly the same manner as in Subsection 1.2. That is, we start by assuming that $y(x)$ is the required stationary path. We then define a neighbouring path

$$\widetilde{y}(x, \varepsilon) = y(x) + \varepsilon\, g(x), \quad \text{where } g(a) = g(b) = 0.$$

Here ε is a small real number, and $g(x)$ is any sufficiently well-behaved function that satisfies $g(a) = g(b) = 0$, so that the neighbouring path $\widetilde{y}$ passes through the same endpoints as y.

Now, the value of the functional along the neighbouring path is given by

$$S[\widetilde{y}] = \int_a^b F(x, \widetilde{y}, \widetilde{y}\,')\,dx, \quad \text{where } \widetilde{y}\,' = \frac{d\widetilde{y}}{x}.$$

For fixed y and g, different values of ε give different values of $S[\widetilde{y}] = S[y + \varepsilon g]$, hence we can think of $S[\widetilde{y}]$ as just a function of ε. But y is defined to be a stationary path, so $S[\widetilde{y}]$ must be stationary at $\varepsilon = 0$; that is,

$$\frac{d}{d\varepsilon}S[\widetilde{y}]\bigg|_{\varepsilon=0} = 0.$$

This is exactly the same argument as in Subsection 1.2, and it leads to the same result, equation (4).

The rate of change of the functional is given by

$$\begin{aligned}\frac{d}{d\varepsilon}S[\widetilde{y}]\bigg|_{\varepsilon=0} &= \left(\frac{d}{d\varepsilon}\int_a^b F(x,\widetilde{y},\widetilde{y}\,')\,dx\right)\bigg|_{\varepsilon=0} \\ &= \int_a^b \left(\frac{d}{d\varepsilon}F(x,\widetilde{y},\widetilde{y}\,')\right)\bigg|_{\varepsilon=0} dx, \end{aligned} \tag{15}$$

where we have used the fact that the integration limits a and b are independent of ε, so that the order of integration and differentiation can be interchanged using Leibniz's integral rule.

Remember that $\widetilde{y}$ and $\widetilde{y}\,'$ both depend on ε, but x does not, so we can apply the chain rule to give

$$\frac{d}{d\varepsilon}F(x,\widetilde{y},\widetilde{y}\,')\bigg|_{\varepsilon=0} = \left(\frac{\partial F}{\partial \widetilde{y}}\frac{d\widetilde{y}}{d\varepsilon} + \frac{\partial F}{\partial \widetilde{y}\,'}\frac{d\widetilde{y}\,'}{d\varepsilon}\right)\bigg|_{\varepsilon=0}.$$

However, $\widetilde{y} = y + \varepsilon g$, hence $\widetilde{y}\,' = y' + \varepsilon g'$, so

$$\frac{d}{d\varepsilon}F(x,\widetilde{y},\widetilde{y}\,')\bigg|_{\varepsilon=0} = \left(\frac{\partial F}{\partial \widetilde{y}}\,g + \frac{\partial F}{\partial \widetilde{y}\,'}\,g'\right)\bigg|_{\varepsilon=0}.$$

Now, $F = F(x,\widetilde{y},\widetilde{y}\,')$ depends on ε, whereas $g = g(x)$ is independent of ε. As $\varepsilon \to 0$, we have $\widetilde{y} \to y$ and $\widetilde{y}\,' \to y'$. Therefore

$$\frac{d}{d\varepsilon}F(x,\widetilde{y},\widetilde{y}\,')\bigg|_{\varepsilon=0} = \frac{\partial F}{\partial y}\,g + \frac{\partial F}{\partial y'}\,g',$$

where now on the right-hand side $F = F(x,y,y')$. The right-hand-side expression is independent of ε.

Substituting this expression into equation (15) gives

$$\frac{d}{d\varepsilon}S[\widetilde{y}]\bigg|_{\varepsilon=0} = \int_a^b \left(g(x)\,\frac{\partial F}{\partial y} + g'(x)\,\frac{\partial F}{\partial y'}\right) dx.$$

The second term in this integral can be rewritten using integration by parts, to give

$$\int_a^b g'(x)\,\frac{\partial F}{\partial y'}\,dx = \left[g(x)\,\frac{\partial F}{\partial y'}\right]_a^b - \int_a^b g(x)\,\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) dx.$$

But $g(a) = g(b) = 0$, so the first term on the right-hand side vanishes, and the rate of change of the functional becomes

$$\begin{aligned}\frac{d}{d\varepsilon} S[\widetilde{y}]\bigg|_{\varepsilon=0} &= \int_a^b \left[g(x)\, \frac{\partial F}{\partial y} - g(x)\, \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] dx \\ &= -\int_a^b g(x) \left[\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y}\right] dx. \end{aligned} \tag{16}$$

If $S[y]$ is stationary, then by definition this integral vanishes for all functions $g(x)$. Because g is arbitrary, this can occur only if

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0, \quad y(a) = A,\ y(b) = B. \tag{17}$$

Deriving the differential equation

You can see how (17) follows from (16) using a 'proof by contradiction', as follows.

Suppose that the bracketed expression on the right of equation (16) is not zero everywhere. Then because $g(x)$ is chosen arbitrarily, we can choose it to have the same sign as the bracketed expression for every value x. Hence for this choice of g, the integrand is positive whenever the bracketed expression is not zero. As a consequence, the integral is positive too, which contradicts the fact that y is a stationary function. Hence the bracketed expression must be zero everywhere, which establishes equation (17).

Equation (17) is known as the *Euler–Lagrange equation*, and is hugely important in mathematical physics. Solutions of the Euler–Lagrange equation are stationary paths of the functional $S[y]$ in equation (14). When applied to a variational problem, the Euler–Lagrange equation usually produces a second-order differential equation, which must be solved subject to the boundary conditions in equation (14) in order to obtain the stationary path.

Note that it is usually much easier to use the Euler–Lagrange equation to find stationary paths than it is to use the type of analysis that we performed in Subsection 1.2.

The Euler–Lagrange equation is the main result of this unit, and it is essential that you remember the equation and how to use it (although you do not need to remember how it is derived).

Euler–Lagrange equation

The functional

$$S[y] = \int_a^b F(x, y, y')\, dx, \quad y(a) = A,\ y(b) = B, \tag{18}$$

has stationary paths that are given by solving the **Euler–Lagrange equation**

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0, \quad y(a) = A,\ y(b) = B. \tag{19}$$

Let us apply the Euler–Lagrange equation in an example and some exercises.

Example 2

Find the stationary path of the distance functional

$$S[y] = \int_a^b \sqrt{1 + y'(x)^2}\, dx, \quad y(a) = A,\ y(b) = B, \tag{20}$$

considered in Section 1 (equation (2)).

Solution

Comparing equations (18) and (20), we see that the integrand is given by

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

To use the Euler–Lagrange equation, we need to calculate $\partial F/\partial y$ and $\partial F/\partial y'$. The first thing to notice is that the integrand depends explicitly only on y', *not* on x or y. Therefore

$$\frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{1 + (y')^2}} \quad \text{and} \quad \frac{\partial F}{\partial y} = 0.$$

Hence the Euler–Lagrange equation (19) becomes

$$\frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) = 0. \tag{21}$$

We could differentiate this to get a second-order differential equation for y, which we could then solve by integration. However, equation (21) is already in a convenient form to be integrated directly.

Integrating both sides of equation (21) with respect to x gives

$$\frac{y'}{\sqrt{1 + (y')^2}} = \alpha, \tag{22}$$

for some constant α. This is identical to equation (6), which is the equation that we derived in Subsection 1.2 for the stationary path. Its solution (equation (7)), was shown to be

$$y = \frac{B - A}{b - a}x + \frac{Ab - Ba}{b - a}.$$

Exercise 7

Consider the functional

$$S[y] = \int_a^b \left((y')^2 + y\right) dx, \quad y(0) = 0,\ y(1) = 2.$$

(a) Show that the stationary paths of this functional must satisfy

$$2y'' - 1 = 0, \quad y(0) = 0,\ y(1) = 2.$$

(b) Solve this differential equation to show that the stationary path is

$$y = \tfrac{1}{4}x^2 + \tfrac{7}{4}x.$$

Exercise 8

(a) Show that the Euler–Lagrange equation for the functional

$$S[y] = \int_0^X \left((y')^2 - y^2\right) dx, \quad y(0) = 0,\ y(X) = 1,$$

where $0 < X < \pi$, is

$$y'' + y = 0.$$

(b) Hence show that the stationary function is $y = \sin x / \sin X$.

Exercise 9

(a) Show that the Euler–Lagrange equation for the functional

$$S[y] = \int_0^1 \left((y')^2 + y^2 + 2xy\right) dx, \quad y(0) = 0,\ y(1) = \alpha,$$

where α is a constant, is

$$y'' - y = x.$$

(b) Hence show that the stationary function is

$$y = \left(\frac{\alpha + 1}{e - e^{-1}}\right)\left(e^x - e^{-x}\right) - x.$$

4.2 Alternative forms of the Euler–Lagrange equation

In Example 2, you saw an example of an integrand

$$F(x, y, y') = \sqrt{1 + y'(x)^2}$$

used to define a functional, which did not depend explicitly on y. In this case, $\partial F/\partial y = 0$, and the Euler–Lagrange equation (19) reduces to

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0.$$

After integrating with respect to x, we obtain

$$\frac{\partial F}{\partial y'} = \text{constant}.$$

(Compare these two equations with equations (21) and (22).) Thus we arrive at the following result.

Special case of the Euler–Lagrange equation

If a functional has the form

$$S[y] = \int_a^b F(x, y')\,dx, \quad y(a) = A,\ y(b) = B, \tag{23}$$

in which the integrand does not depend explicitly on y, then the Euler–Lagrange equation can be integrated to give

$$\frac{\partial F}{\partial y'} = \text{constant}.$$

If, on the other hand, the integrand does not depend explicitly on x, so that the functional has the form

$$S[y] = \int_a^b F(y, y')\,dx, \quad y(a) = A,\ y(b) = B,$$

then it can be shown that the Euler–Lagrange equation (19) reduces to

$$y' \frac{\partial F}{\partial y'} - F = c, \quad y(a) = A,\ y(b) = B,$$

where c is a constant determined by the boundary conditions. We will derive this equation in Exercise 12. The expression on the left-hand side of the equation is called the *first integral* of the Euler–Lagrange equation.

First integral of the Euler–Lagrange equation

If a functional has the form

$$S[y] = \int_a^b F(y, y')\,dx, \quad y(a) = A,\ y(b) = B, \tag{24}$$

where the integrand does not depend explicitly on x, then the **first integral** of the Euler–Lagrange equation is

$$y' \frac{\partial F}{\partial y'} - F = c, \quad y(a) = A,\ y(b) = B, \tag{25}$$

where c is a constant determined by the boundary conditions.

The stationary path of the functional S is determined by solving this first-order differential equation.

Equations (23) and (25) are important because they both give rise to first-order differential equations, whereas the most general form of the Euler–Lagrange equation gives rise to a second-order differential equation – and first-order equations are *usually* easier to solve than second-order equations.

Exercise 10

Consider the functional

$$S[y] = \int_1^2 (\ln y' + y)\, dx, \quad y(1) = \ln 2,\ y(2) = \ln 3.$$

(a) Show that the first integral of the Euler–Lagrange equation is

$$\frac{dy}{dx} = \alpha e^{-y}, \quad y(1) = \ln 2,\ y(2) = \ln 3,$$

for some constant α.

(b) Hence show that the stationary function is $y = \ln(x+1)$.

Exercise 11

Consider the functional

$$S[y] = \int_{-1}^1 \sqrt{y(1+(y')^2)}\, dx, \quad y(-1) = y(1) = A,$$

where $A \geq 1$.

(a) Show that the first integral of the Euler–Lagrange equation for this functional is

$$y' = \frac{1}{k}\sqrt{y - k^2}, \quad y(-1) = y(1) = A.$$

(b) Hence show that the functional is stationary on the paths

$$y(x) = h + \frac{x^2}{4h}, \quad \text{where } h = \tfrac{1}{2}\big(A \pm \sqrt{A^2 - 1}\big).$$

Exercise 12

(a) Suppose that $F(y, y')$ does not depend explicitly on x, so that $\partial F/\partial x = 0$. Show that

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

(*Hint*: Evaluate the derivative on the left-hand side.)

(b) Hence show that equation (25) is true if either (i) $y(x)$ satisfies the Euler–Lagrange equation, or (ii) $y(x)$ is constant.

5 Applications of the Euler–Lagrange equation

In this section we consider some applications of the Euler–Lagrange equation to physical problems. We begin with a discussion of *Fermat's principle*, which describes the trajectories of light rays. This topic is the most important one of this section because it forms a useful stepping stone to the next unit, on Lagrangian mechanics.

After Fermat's principle, we revisit two of the examples of Section 2 – the brachistochrone and the soap film. These are interesting applications that expose you to the sorts of method used to construct functionals, and they exhibit the type of behaviour that can be modelled using the calculus of variations. However, solving the Euler–Lagrange equation for these two problems is beyond what we would expect from you in assessment of this module. So do not worry if some of the details of Subsections 5.2 and 5.3 seem more advanced than the rest of this unit. You will not be assessed on the fine details of these topics. However, the examples and exercises distributed throughout this section *are* typical of the sorts of question on which you will be assessed. It is therefore very important that you follow the examples and complete the exercises.

5.1 Fermat's principle

In most circumstances, light can be considered to travel along lines from a source (such as a light bulb) to an observer (such as a human eye); these lines are called **light rays**, or just **rays**, and are often straight lines. This is why most shadows have distinct edges, and why eclipses of the Sun are so spectacular.

A simple experiment with light rays involves reflecting light in a plane mirror, as shown in Figure 12. In the diagram there is a fixed point source of light S, which emits light rays in all directions. One light ray is reflected from the mirror at R and travels to an observer fixed at O, as shown in the diagram. Labelled in the figure are the angle θ_1 between the line SR and the vertical line through R, and the angle θ_2 between the line RO and the vertical line through R. (The plane of the mirror is perpendicular to the page, and it is assumed that the plane SRO is in the page.)

We wish to determine the location of R. The key to this, obtained from experiment, is to observe that for light rays reflected in this way, the incoming angle θ_1, called the **angle of incidence**, is equal to the outgoing angle θ_2, called the **angle of reflection**. Using the fact that $\theta_1 = \theta_2$, we can determine the x-coordinate of R using elementary geometry. (In brief, because $\theta_1 = \theta_2$, the two right-angled triangles with hypotenuses SR and OR are similar, from which it follows that $x/h_1 = (d-x)/h_2$. Rearranging this equation gives $x = dh_1/(h_1 + h_2)$. We will not use this exact formula for x again.)

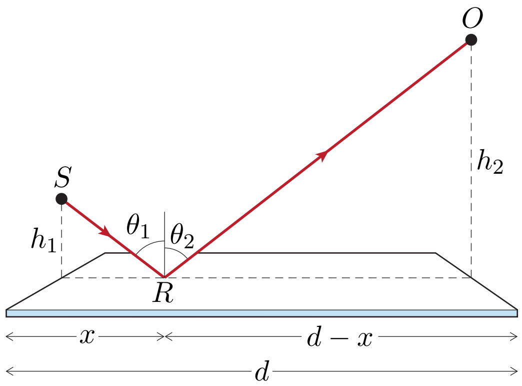


Figure 12 A light ray emitted from S, reflected at R, and observed at O

This observation about the two angles has an interesting consequence. Consider any path from S to R to O of the type shown in Figure 12 (not necessarily the path of a light ray; θ_1 and θ_2 may differ). By applying Pythagoras' theorem to the triangles, we can see that the total length of the path is

$$f(x) = \sqrt{x^2 + h_1^2} + \sqrt{(d-x)^2 + h_2^2}.$$

It turns out (see Exercise 13) that this function of x has a minimum when $\theta_1 = \theta_2$; so the light ray that travels from S to O via reflection in the mirror minimises the distance that it travels.

This result was generalised by the French mathematician Pierre de Fermat (1601–1665) into what is now known as the *principle of least time*.

The **principle of least time** states that the path taken between two points by a ray of light is the path that can be traversed in the least time.

(Note that this principle is not *quite* correct; an amended statement will be given shortly.)

This principle means that light will always travel in straight lines when it passes through a single uniform medium like air, because it has a constant speed, so the shortest distance is always the fastest time of passage. However, light has different speeds in different media: for example, in a vacuum the speed of light is $c \simeq 2.9 \times 10^8\,\mathrm{m\,s^{-1}}$, whereas in water it is $c_{\mathrm{water}} \simeq 2.25 \times 10^8\,\mathrm{m\,s^{-1}}$. (Note that the letter c is conventionally used to denote the speed of light in vacuum. Take care not to get this mixed up with any constants of integration that you introduce!) Generally, if the speed of light in a uniform medium is c_{m}, then we define the **refractive index** of the medium to be the ratio $n = c/c_{\mathrm{m}}$; so the refractive index of a medium is just another way of characterising the speed of light that passes through it.

Consider now a general (possibly non-uniform) medium in the Cartesian plane, with refractive index $n(x,y)$ at the point (x,y). Then the speed of light will vary in space according to the formula $c_{\mathrm{m}}(x,y) = c/n(x,y)$, and the time taken for light to travel along an infinitesimal line segment of length δs is

$$\delta t = \frac{\delta s}{c_{\mathrm{m}}(x,y)} = \frac{1}{c}\, n(x,y)\, \delta s.$$

However, in Subsection 1.1 we showed that along a path $y = y(x)$, we have $\delta s = \sqrt{1 + y'(x)^2}\, \delta x$ (see equation (1)). Therefore

$$\delta t = \frac{1}{c}\, n(x,y) \sqrt{1 + y'(x)^2}\, \delta x.$$

So for light moving in the Cartesian plane, in a medium with refractive index $n(x,y)$, with the source at the point (a,A) and observer at the point (b,B), the time of passage T along a path $y(x)$ joining these points is

$$T[y] = \frac{1}{c} \int_a^b n(x,y) \sqrt{1 + (y')^2}\, dx, \quad y(a) = A,\ y(b) = B. \tag{26}$$

In the context of the calculus of variations, the principle of least time states that the path taken will minimise this functional. In fact, this is not quite correct: light rays actually take paths that make this functional stationary; that is, light rays follow paths that are stationary paths of equation (26). Normally, such paths *do* minimise the time of passage, but there are situations in which they also maximise it. We call our revised principle *Fermat's principle*; it supersedes the principle of least time stated earlier.

Fermat's principle says that the path taken between two points by a ray of light is a stationary path of the time functional (26).

In other texts, the phrase 'principle of least time' is used for the statement that we call Fermat's principle. Here we reserve the phrase 'Fermat's principle' for the correct statement about the path travelled by light rays.

For light travelling through a uniform material (i.e. $n(x,y) = n$, for some constant n and all points (x,y)), the functional (26) is proportional to the distance functional of Subsection 1.1. Therefore the stationary paths are straight lines. For a general non-uniform medium, in which the refractive index depends on the position, light rays follow curved paths. *Mirages* are one consequence of a position-dependent refractive index, where light passes through layers of air of differing temperature and therefore density, leading to distorted images – see Figure 13.

Figure 13 Sunset mirage

Snell's law of refraction

Fermat's principle can be used to show that for light reflected in a mirror, the angle of incidence equals the angle of reflection. For light crossing the boundary between two media, it gives **Snell's law**, named after the Dutch astronomer and mathematician Willebrord Snellius (1580–1626), which says that

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{c_1}{c_2},$$

where α_1 and α_2 are the angles between the ray and the normal to the boundary, and c_1 and c_2 are the speeds of light in the media, as shown in Figure 14.

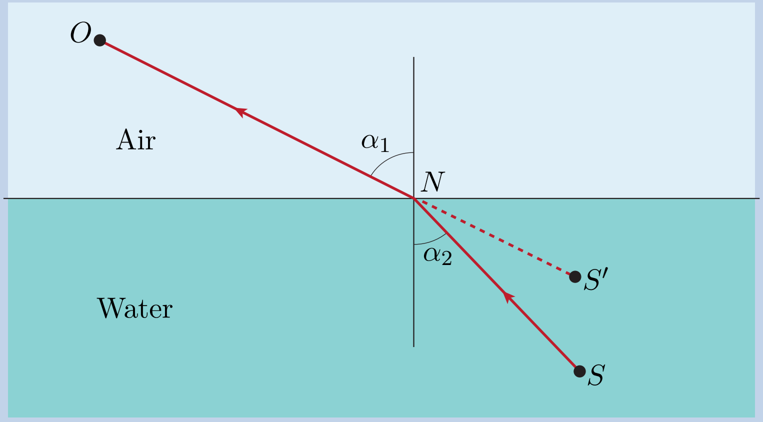


Figure 14 A ray of light is refracted as it passes from water to air

In water the speed of light is approximately $c_2 = c_1/1.3$, where c_1 is the speed of light in air. It follows that $1.3\sin\alpha_2 = \sin\alpha_1$, so $\alpha_1 > \alpha_2$. In Figure 14, the light from an object S in water travels to the observer O in the air along the two straight lines SN and NO. The observer perceives the object to be at S', on the straight line through ON. This explains why, for instance, a stick partly submerged in water appears bent.

Intuitively, we can see why light bends in this way: its speed is greater in air than in water, so to minimise the time of travel, it needs to extend its path in air and reduce its path in water.

The situation is similar to that of a lifeguard on a beach, who wants to reach a swimmer in distress in the fastest possible time. Since the lifeguard can run faster than she can swim, she must extend her path on the sand and reduce her path in the water, as shown in Figure 15.

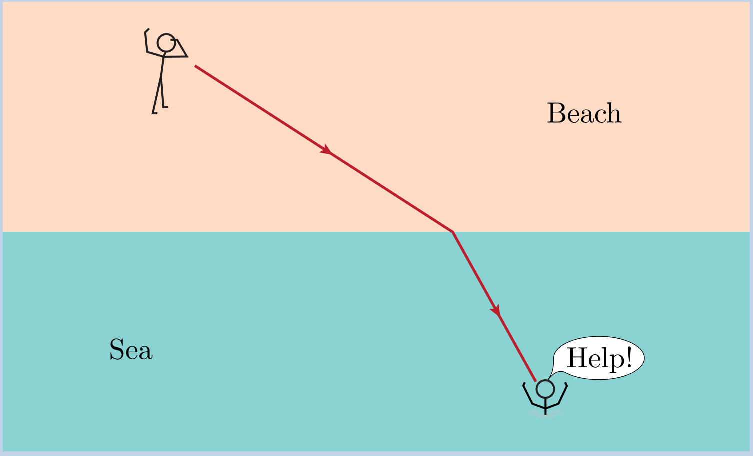


Figure 15 Path taken by a lifeguard to reach a distressed swimmer

Connection to Lagrangian mechanics

Equation (26) is sometimes written in the form

$$T[y] = \int_a^b L(x, y, y')\, dx, \quad y(a) = A,\ y(b) = B, \tag{27}$$

where

$$L(x, y, y') = \frac{1}{c}\, n(x, y)\sqrt{1 + (y')^2}$$

is called the **optical Lagrangian** and T is called the **optical action**. Fermat's principle tells us that light rays passing between (a, A) and (b, B) travel along paths that make the optical action stationary.

This is analogous to what we will find in the next unit. There we will demonstrate that the equations of motion of mechanical objects can be obtained by finding the stationary paths of an *action* like (27), but with a different form for the *Lagrangian* L, and where the independent variable is time t instead of distance x.

Example 3

Consider a layer of transparent medium with refractive index $n(x, y) = 4\sqrt{y}$ that is located in the part of the plane given by $\frac{1}{8} \leq y \leq 2$. Find the path of a light ray from a source at $\left(-1, \frac{5}{4}\right)$ to an observer at $\left(1, \frac{5}{4}\right)$.

Solution

By Fermat's principle, the light ray will be a stationary path of the functional

$$T[y] = \frac{4}{c}\int_{-1}^{1} \sqrt{y(1 + (y')^2)}\, dx, \quad y(-1) = \tfrac{5}{4},\ y(1) = \tfrac{5}{4}.$$

Apart from the factor $4/c$, this is exactly the functional of Exercise 11, with the same boundary conditions if $A = \frac{5}{4}$.

From the solution to Exercise 11(b), the functional is stationary on the two paths

$$y(x) = h + \frac{x^2}{4h}, \quad \text{where } h = \tfrac{1}{2}\big(A \pm \sqrt{A^2 - 1}\big).$$

As $A = \frac{5}{4}$, we see that $h = 1$ or $h = \frac{1}{4}$, so there are two solutions:

$$y_+(x) = 1 + \tfrac{1}{4}x^2 \quad \text{and} \quad y_-(x) = \tfrac{1}{4} + x^2.$$

The paths described by y_+ and y_- are shown in Figure 16. The observer at $\left(1, \frac{5}{4}\right)$ sees two light rays coming from different directions.

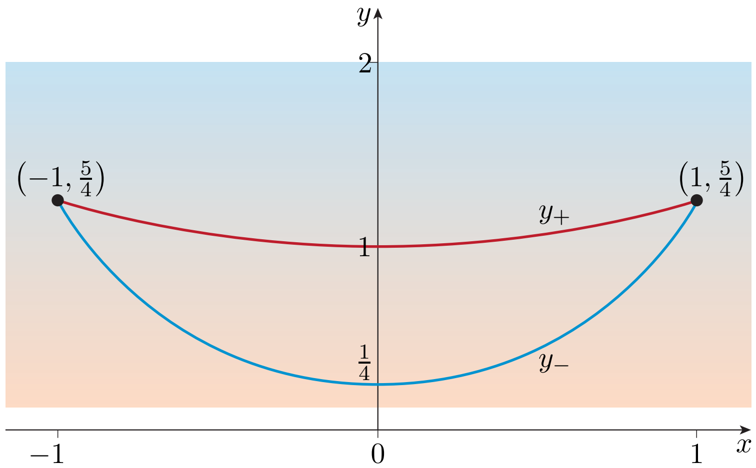


Figure 16 Graph of solutions

Note that the paths of the light rays depend on the refractive index n, but they do not depend on c, the explicit value of the speed of light in a vacuum.

Here are some exercises for you to try.

Exercise 13

Show that the function

$$f(x) = \sqrt{x^2 + h_1^2} + \sqrt{(d - x)^2 + h_2^2},$$

where h_1, h_2, x and d represent the distance measurements of Figure 12, is stationary when $\theta_1 = \theta_2$, where

$$\sin\theta_1 = \frac{x}{\sqrt{x^2 + h_1^2}} \quad \text{and} \quad \sin\theta_2 = \frac{d - x}{\sqrt{(d - x)^2 + h_2^2}}.$$

Show that at this stationary value, $f(x)$ has a minimum.

Exercise 14

Consider a layer of transparent medium with refractive index given by $n(x,y) = 2/y$ that is located in the part of the plane given by $\frac{1}{2} \leq y \leq 2$.

(a) Use Fermat's principle to show that the path of a light ray located within this medium is described by the differential equation

$$\frac{dy}{dx} = \frac{1}{ky}\sqrt{1-k^2y^2} \quad \text{for } \tfrac{1}{2} \leq y \leq 2,$$

where k is a constant.

(b) Find the path of a light ray from source at $(-1,1)$ to observer at $(1,1)$.

5.2 Revisiting the brachistochrone

In Subsection 2.1, you learned about the problem of finding the shape of a wire joining two given points, such that a bead starting from rest from one point will slide down the wire under gravity to the other point, without friction, in the shortest time. Here you will first see how to derive the functional for this system, and then learn how to solve it.

Deriving the functional of the brachistochrone problem

Let us formulate the functional for the brachistochrone problem by obtaining an expression for the time of passage between given points $P_a = (a,A)$ and $P_b = (b,B)$ along a curve $y(x)$, using the principle of conservation of energy. Such arguments involving energy are possible only because we have assumed that the bead slides without friction.

We use a coordinate system with the y-axis oriented vertically downwards. Let us suppose that the particle starts *from rest* at the origin; that is, $a = 0$ and $A = 0$. To simplify the problem further, we assume that the end of the wire is a unit distance below the starting point; that is, $P_b = (b,1)$, where b is positive and is allowed to vary. A typical curve $y(x)$ is shown in Figure 17. Labelled in that figure is the distance $s(x)$ from the origin to a point $P = (x, y(x))$ on the curve.

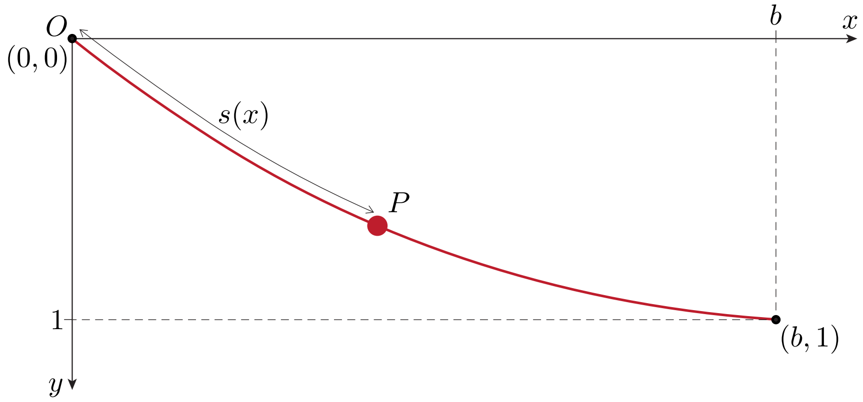


Figure 17 A point P on the curve $y(x)$

The speed of the bead at the point $P = (x, y(x))$ is $v = ds/dt$. Let m denote the mass of the bead. Then the kinetic energy of the bead at P is $\frac{1}{2}mv^2$, and the potential energy of the bead at that point is $-mgy$. (The sign of the potential energy is negative because the y-axis points downwards; we therefore expect potential energy to decrease as y increases.)

The total energy E of the bead is the sum of its kinetic and potential energies:

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

Initially, the value of E is 0, because the initial values of v and y are 0. Moreover, because there is no friction, energy is conserved, so E is always 0; that is,

$$\tfrac{1}{2}mv^2 - mgy = 0. \tag{28}$$

As we saw earlier (Figure 3), small changes in s are given by $\delta s^2 = \delta x^2 + \delta y^2$, so

$$\begin{aligned}\left(\frac{ds}{dt}\right)^2 &= \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \\ &= \left(\frac{dx}{dt}\right)^2 \left(1 + \left(\frac{dy}{dt}\right)^2 \bigg/ \left(\frac{dx}{dt}\right)^2\right) \\ &= \left(\frac{dx}{dt}\right)^2 \left(1 + (y')^2\right).\end{aligned}$$

Here we have used the fact that

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt},$$

by the chain rule. As $v = ds/dt$, we can write equation (28) as

$$\tfrac{1}{2}m\left(\frac{dx}{dt}\right)^2 \left(1 + (y')^2\right) - mgy = 0.$$

Rearranging this gives

$$\left(\frac{dx}{dt}\right)^2 = \frac{2gy}{1 + (y')^2}.$$

Now, the time of passage from $x = 0$ (at time $t = 0$) to $x = b$ (at time $t = T$) is given by the integral

$$T = \int_0^T dt = \int_0^b \frac{1}{dx/dt}\, dx = \int_0^b \sqrt{\frac{1 + (y')^2}{2gy}}\, dx.$$

Therefore the functional for the time taken for a bead to slide down the wire with shape $y(x)$, starting from rest at the origin, is

$$T[y] = \frac{1}{\sqrt{2g}} \int_0^b \sqrt{\frac{1 + (y')^2}{y}}\, dx, \quad y(0) = 0,\ y(b) = 1. \tag{29}$$

Exercise 15

Use equation (29) to show that the time taken for a particle of mass m to slide down the curve $y = x/b$ from rest at the origin to the point $(b, 1)$ is

$$T = 2\sqrt{\frac{b^2+1}{2g}}.$$

Minimising the functional of the brachistochrone problem

To determine the shape of the brachistochrone, we need to minimise the functional (29). The integrand

$$F(y, y') = \frac{1}{\sqrt{2g}}\sqrt{\frac{1+(y')^2}{y}}$$

of this functional is independent of x, so we can use the first integral of the Euler–Lagrange equation (25):

$$y'\,\frac{\partial F}{\partial y'} - F = \text{constant}.$$

We obtain

$$\frac{1}{\sqrt{2g}}\,\frac{(y')^2}{\sqrt{y(1+(y')^2)}} - \frac{1}{\sqrt{2g}}\sqrt{\frac{1+(y')^2}{y}} = \text{constant}.$$

Simplifying the left-hand side of this equation, and multiplying by -1, gives

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

where c is a constant. (We have absorbed the factor $\sqrt{2g}$ into the constant.) Squaring both sides and rearranging gives $y\left(1+(y')^2\right) = c^2$, from which we get

$$\frac{dy}{dx} = \pm\sqrt{\frac{c^2}{y} - 1} = \pm\sqrt{\frac{c^2-y}{y}}.$$

This first-order differential equation is separable and can be solved. First, however, note that because the y-axis is vertically downwards, we expect the solution $y(x)$ to increase as x increases, so we take the positive sign in this equation. Separating the variables now gives

$$\int dx = \int \sqrt{\frac{y}{c^2-y}}\,dy.$$

Of course, the integral on the left-hand side is just x (plus a constant, which we absorb into the constant from the right-hand side).

To solve the integral on the right-hand side, we substitute $y = c^2 \sin^2 \phi$, so that $dy = 2c^2 \sin\phi \cos\phi \, d\phi$, to obtain

$$\int \sqrt{\frac{y}{c^2 - y}} \, dy = 2c^2 \int \sin\phi \cos\phi \sqrt{\frac{c^2 \sin^2 \phi}{c^2 - c^2 \sin^2 \phi}} \, d\phi.$$

Using the trigonometric identities $1 - \sin^2 \phi = \cos^2 \phi$ and $\sin^2 \phi = \frac{1}{2}(1 - \cos 2\phi)$, we obtain

$$\begin{aligned}\int \sqrt{\frac{y}{c^2 - y}} \, dy &= 2c^2 \int \sin\phi \cos\phi \sqrt{\frac{\sin^2 \phi}{\cos^2 \phi}} \, d\phi \\ &= 2c^2 \int \sin^2 \phi \, d\phi \\ &= c^2 \int (1 - \cos 2\phi) \, d\phi \\ &= \tfrac{1}{2} c^2 (2\phi - \sin 2\phi) + d,\end{aligned}$$

where d is a constant.

In summary, we have

$$x = \tfrac{1}{2} c^2 (2\phi - \sin 2\phi) + d \quad \text{and} \quad y = c^2 \sin^2 \phi = \tfrac{1}{2} c^2 (1 - \cos 2\phi),$$

which are solutions of the functional in parametric form (where the parameter is ϕ). When $\phi = 0$, these equations give $x = d$ when $y = 0$. However, we chose our axes such that $x = 0$ when $y = 0$; hence $d = 0$, and the equations become

$$x = \tfrac{1}{2} c^2 (2\phi - \sin 2\phi) \quad \text{and} \quad y = \tfrac{1}{2} c^2 (1 - \cos 2\phi). \tag{30}$$

The curve described by these equations passes through the origin, and the constant c is determined by the fact that the curve also passes through $(b, 1)$. Suppose that $\phi = \phi_b$ when $(x, y) = (b, 1)$. Then

$$b = \tfrac{1}{2} c^2 (2\phi_b - \sin 2\phi_b) \quad \text{and} \quad 1 = \tfrac{1}{2} c^2 (1 - \cos 2\phi_b),$$

so $\frac{1}{2} c^2 = 1/(1 - \cos 2\phi_b)$. Therefore our parametric equations are

$$x = \frac{2\phi - \sin 2\phi}{1 - \cos 2\phi_b} \quad \text{and} \quad y = \frac{1 - \cos 2\phi}{1 - \cos 2\phi_b} \quad \text{for } 0 \leq \phi \leq \phi_b, \tag{31}$$

where ϕ_b is given by

$$b = \frac{2\phi_b - \sin 2\phi_b}{1 - \cos 2\phi_b}. \tag{32}$$

Clearly, if $\phi = 0$, then $(x, y) = (0, 0)$, and if $\phi = \phi_b$, then $(x, y) = (b, 1)$.

Equation (32) cannot be solved analytically for ϕ_b (there is no closed formula for ϕ_b), and instead we must solve it numerically. Figure 18 shows a graph of b against ϕ_b. The graph starts at the origin, and b increases as ϕ_b increases. You can see from the graph that $b \to \infty$ as $\phi_b \to \pi$, a fact easily verified using equation (32). The graph shows us that for every value of b, a unique value of ϕ_b can be found. You are asked to reproduce this graph in Exercise 18.

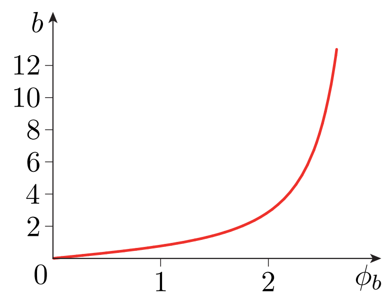


Figure 18 Graph of b as a function of ϕ_b

To solve the brachistochrone problem for any value b, we must solve equation (32) for ϕ_b, then substitute the result into the parametric equations (31). The result can be plotted using suitable computer software. Figure 19 shows graphs of the stationary paths for various values of b; all the curves start at $(0,0)$ and end at $(b,1)$.

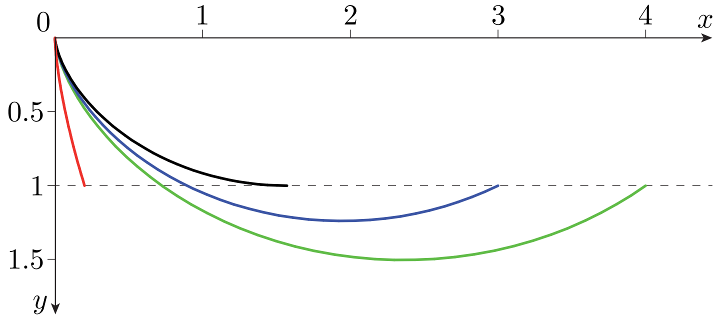


Figure 19 Graphs of the stationary paths joining the points $(0,0)$ and $(b,1)$ for $b=0.2$ (red), $b=\pi/2$ (black), $b=3$ (blue) and $b=4$ (green)

From the figure, we see that for small values of b, the stationary path is close to that of a straight line, as you might expect. In fact, if $b<\pi/2$, then the stationary path crosses the $y=1$ line just once, at the terminal point. For $b>\pi/2$, the stationary path crosses the $y=1$ line twice. The critical value of b, where the stationary path is tangential to the line $y=1$ at its terminal point, occurs at $b=\pi/2$ (see Exercise 16(c) for a derivation of this result).

Cycloid

A **cycloid** is the curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line without slippage. Figure 20 illustrates a cycloid.

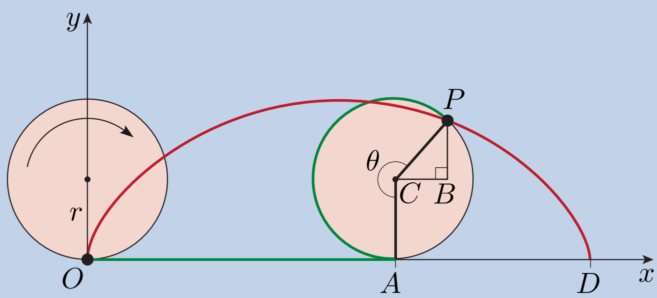


Figure 20 Cycloid (in red) traced out by a point on the rim of a rolling wheel

In this figure, a circle of radius r rolls along the x-axis, starting with its centre on the y-axis. Fix attention on the point P attached to the circle, initially at the origin O. As the circle rolls, P traces out the cycloid through O, P and D.

The cycloid has been studied by many mathematicians from the time of Galileo (1564–1642), and was the cause of so many controversies and quarrels in the 17th century that it became known as ‘the Helen of geometers’ (a reference to ‘Helen of Troy’ from Greek mythology). Galileo named the cycloid but had limited understanding of the mathematics behind it. He did, however, suggest that it would make an attractive arch for a bridge. This suggestion was implemented in 1764 with the building by James Essex (1722–1784) of Trinity College Bridge, in the grounds of Trinity College, over the river Cam in Cambridge, which has three cycloidal arches. The cycloid is still occasionally used in architecture today. Other notable mathematicians to have studied the cycloid include René Descartes (1596–1650), Pierre de Fermat (1601–1665), Blaise Pascal (1623–1662), Christiaan Huygens (1629–1695) and Christopher Wren (1632–1723), the architect of St Paul’s Cathedral.

The equation of the cycloid is obtained by finding the coordinates of P. In Figure 20, after the circle has rolled through an angle θ, the length of the longer circular arc between P and A is $r\theta$. Because there is no slipping, $OA = r\theta$ also, so C has coordinates $(r\theta, r)$. Next, by applying trigonometry to the triangle with vertices P, B and C, we see that $PB = -r\cos\theta$ and $BC = -r\sin\theta$, so P has coordinates

$$x = r(\theta - \sin\theta) \quad \text{and} \quad y = r(1-\cos\theta),$$

which together give the parametric equation of the cycloid. These are identical to equation (30) if we make the transformation $\theta = 2\phi$ and $r = c^2/2$. So the shape of the brachistochrone has a simple geometric interpretation: it is a segment of a cycloid.

Exercise 16

(a) Show that the gradient of the brachistochrone described by equations (31) and (32) is

$$\frac{dy}{dx} = \frac{\sin 2\phi}{1-\cos 2\phi} = \frac{1}{\tan\phi}.$$

(*Hint*: Use the identities $\sin 2\phi = 2\sin\phi\cos\phi$ and $\cos 2\phi = 1 - 2\sin^2\phi$.)

(b) Deduce that at the origin $(0,0)$, the brachistochrone hangs vertically downwards.

(c) Show that the critical brachistochrone, which is the brachistochrone with parameter b that is tangent to the line $y = 1$ at its terminal point (the black curve in Figure 19), satisfies $b = \pi/2$.

Exercise 17

Consider the functional

$$S[y] = \int_0^1 \sqrt{1 + x + (y')^2}\, dx, \quad y(0) = 0, \; y(1) = 1.$$

(a) Show that the function $y(x)$ defined by the relation

$$y' = c\sqrt{1 + x + (y')^2},$$

where c is a constant, makes $S[y]$ stationary.

(b) By expressing y' in terms of x, solve this equation to obtain

$$y = \frac{(1 + x)^{3/2} - 1}{2^{3/2} - 1}.$$

Exercise 18 *Computing – optional*

(a) Use computer software and equation (32) to create a plot of b against ϕ_b similar to Figure 18.

(b) Modify the code given in the section 'Example of plotting a brachistochrone' of the *Computer Exploration Worksheet* to create a plot of four brachistochrones with parameters $b = 0.2, \pi/2, 3, 4$, similar to Figure 19.

5.3 Revisiting minimal surfaces of revolution

Minimal surfaces of revolution were briefly discussed in Subsection 2.2. The problem is to find the function $y(x)$ with given endpoints $y(a) = A$ and $y(b) = B$, where $A, B \geq 0$, such that the area of the surface formed by rotating the curve $y(x)$ about the x-axis is minimised. You saw that to a good approximation, this surface is the same as that of a soap film suspended between a pair of wire hoops that lie in planes perpendicular to the x-axis, with radii A and B, and centres $(a, 0)$ and $(b, 0)$, respectively.

This subsection has several parts. First we derive the functional $S[y]$ for surface area. Then we apply the Euler–Lagrange equation to obtain the differential equation that a function $y(x)$ must satisfy to make the functional stationary. Next we solve this differential equation in a simple case, discovering that even this special case has subtleties. Finally, we discuss the physical interpretation of our results.

Deriving the surface of revolution functional

Figure 21(a) shows the surface of revolution of a curve $y(x)$ that satisfies $y(a) = A$ and $y(b) = B$. To obtain an expression for the area of this surface, we first find the area of a thin disc of width δx, such as that shown in Figure 21(b). The small segment of length δs can be approximated by a straight line provided that δx is sufficiently small, so using the analysis of Subsection 1.1 (see equation (1)), we obtain

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

The area δS traced out by this segment as it rotates about the x-axis is

$$\delta S = 2\pi\, y(x)\, \delta s = 2\pi\, y(x)\, \sqrt{1 + y'(x)^2}\, \delta x.$$

Hence the area of the whole surface from $x = a$ to $x = b$ is

$$S[y] = 2\pi \int_a^b y(x)\, \sqrt{1 + y'(x)^2}\, dx, \quad y(a) = A,\ y(b) = B, \tag{33}$$

where $A, B \geq 0$.

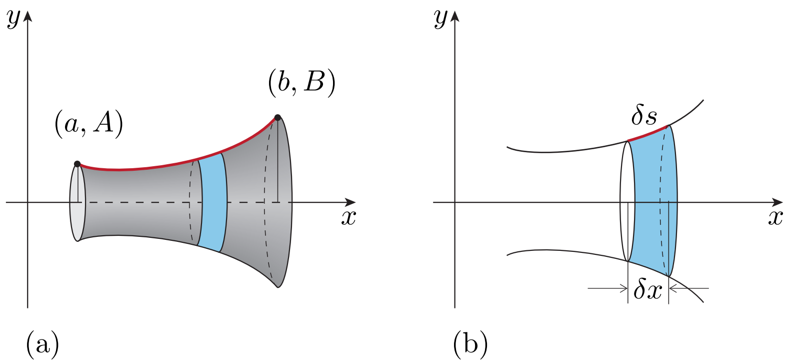


Figure 21 (a) Surface of revolution; (b) a narrow cross-section of that surface

Exercise 19

(a) Show that the equation of the straight line joining the points $(0, 0)$ and (b, B) is

$$y = \frac{B}{b} x.$$

(b) Use equation (33) to show that the area of the cone (without a base) shown in Figure 22 is

$$S = \pi B \sqrt{b^2 + B^2}.$$

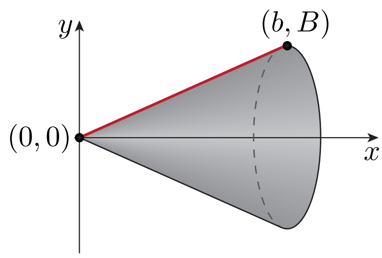


Figure 22 Cone of height b and base radius B, base not included

Applying the Euler–Lagrange equation to the surface of revolution functional

The integrand of the functional (33) is $2\pi\, F(y, y')$, where

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

This function does not depend explicitly on x, so we can use the first integral of the Euler–Lagrange equation. As

$$\frac{\partial F}{\partial y'} = \frac{yy'}{\sqrt{1 + (y')^2}},$$

we see that

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

Applying the first integral of the Euler–Lagrange equation, equation (25), gives

$$\frac{y}{\sqrt{1 + (y')^2}} = y_0, \quad y(a) = A, \; y(b) = B,$$

for some constant y_0 (this constant absorbs the factor 2π that we suppressed earlier). By squaring both sides of this equation and rearranging, we get $y^2/y_0^2 = 1 + (y')^2$, from which we obtain the first-order differential equation

$$\frac{dy}{dx} = \frac{\sqrt{y^2 - y_0^2}}{y_0}, \quad y(a) = A, \; y(b) = B. \tag{34}$$

(Note that the constant y_0, which could be positive or negative, but not zero, absorbs the factor ± 1 that we obtain by taking square roots.) The solutions of equation (34) are stationary paths of the functional (33).

Solving the differential equation for the minimal surface of revolution

The differential equation (34) can be solved by separating variables. The solution involves the hyperbolic cosine function (also known as cosh), but it is not necessary to know many properties of hyperbolic sines or cosines to follow the subsequent analysis. However, if you have not come across these functions previously, then you might like to read the following box, which contains a brief summary of their properties, before continuing.

Brief summary of hyperbolic functions

Recall that Euler's formula $e^{ix} = \cos x + i \sin x$ enables us to express sines and cosines in terms of complex exponentials:

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}) \quad \text{and} \quad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix}).$$

The **hyperbolic sine function**, also called **sinh** (pronounced 'shine', or sometimes 'sinch'), and the **hyperbolic cosine function**, also called **cosh** (pronounced 'cosh'), are defined in analogy to these formulas, but with real exponentials:

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}) \quad \text{and} \quad \sinh x = \tfrac{1}{2}(e^x - e^{-x}).$$

Using these definitions, we can show that hyperbolic sines and cosines have many similar (although not identical) properties to ordinary sines and cosines. For example,

$$\frac{d}{dx}\cosh x = \sinh x \quad \text{and} \quad \frac{d}{dx}\sinh x = \cosh x.$$

Their graphs, however, look quite different, as we can see from Figure 23(a).

The inverses of the hyperbolic functions can be shown to be

$$\begin{aligned} \operatorname{arccosh} x &= \ln(x + \sqrt{x^2 - 1}), \quad x \geq 1, \\ \operatorname{arcsinh} x &= \ln(x + \sqrt{x^2 + 1}). \end{aligned}$$

The graphs of the inverse hyperbolic functions are shown in Figure 23(b).

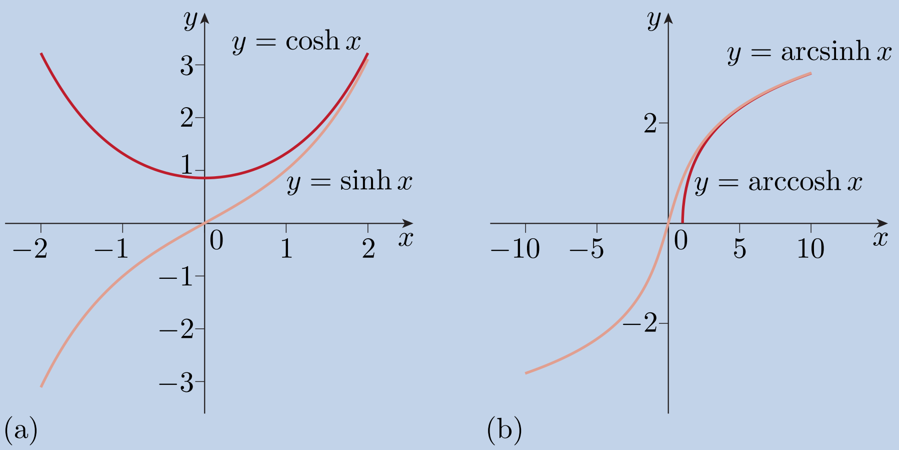


Figure 23 Graphs of: (a) cosh and sinh; (b) arccosh and arcsinh

Separating the variables in equation (34), we obtain

$$\int \frac{dy}{\sqrt{y^2 - y_0^2}} = \frac{1}{y_0}\int dx.$$

Then, referring to the table of standard integrals in the Handbook, we find that

$$\operatorname{arccosh}\left(\frac{y}{y_0}\right) = \frac{x+k}{y_0},$$

where k is a constant.

Rearranging, we obtain

$$y = y_0 \cosh\left(\frac{x+k}{y_0}\right). \tag{35}$$

The constants y_0 and k can be determined using the boundary conditions $y(a) = A$ and $y(b) = B$. We will see that, depending on the values of a, A, b and B, suitable solutions of the minimal surface of revolution problem do not always exist, and that when they do, it may be necessary to carry out further work in order to determine the nature of the stationary point.

Solution of the minimal surface of revolution problem in a special case

In order to understand the nature of the solution (35), let us consider a simple physical example – a soap film suspended between two identical circular hoops of radius 1, centred at $(\pm L/2, 0)$, and each lying in a plane perpendicular to the x-axis (see Figure 24). In this case, $a = -L/2$, $b = L/2$ and $A = B = 1$.

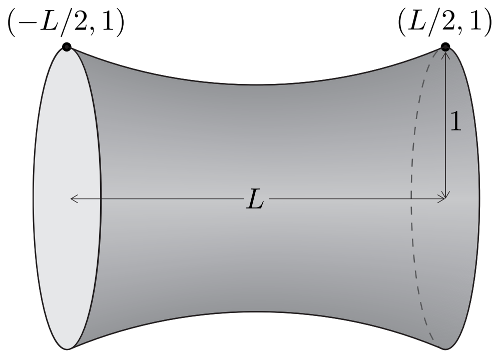


Figure 24 Symmetric soap film

Let us use physical intuition to determine one of the constants. Because the rings are the same size and placed symmetrically about the y-axis, $y(x)$ must be an even function, that is, $y(x) = y(-x)$. However, because cosh is also an even function, there is only one way that solution (35) can be symmetric – namely, if $k = 0$ (the same result is obtained using algebra in Exercise 20). Hence the solution for this symmetric case is

$$y = y_0 \cosh\left(\frac{x}{y_0}\right), \quad y(-L/2) = 1, \; y(L/2) = 1.$$

Setting $x = 0$ gives $y(0) = y_0$, so the constant y_0 is simply the value of y at the origin. Applying the boundary conditions gives $y_0 \cosh(L/2y_0) = 1$, so our solution becomes

$$y = y_0 \cosh\left(\frac{x}{y_0}\right), \quad \text{where } y_0 \cosh\left(\frac{L}{2y_0}\right) = 1. \tag{36}$$

The equation $y_0 \cosh(L/2y_0) = 1$ cannot be rearranged to give a closed formula for y_0; instead, we must solve the equation numerically on a computer. Figure 25 shows a plot of y_0 against L. You are asked to reproduce this graph in Exercise 22.

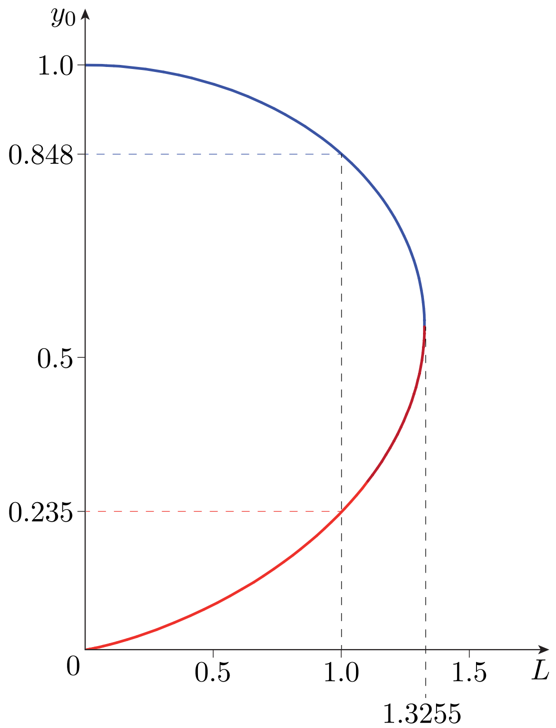


Figure 25 Graph of y_0 as a function of L

From the graph, you can see that for $L > 1.3255$ (approximately) there is no value of y_0 that satisfies the equation $y_0 \cosh(L/2y_0) = 1$, whereas for $L < 1.3255$ there are two solutions of this equation, corresponding to the red and blue parts of the curve. Let us examine each of these cases in turn.

Case 1: $L < 1.3255$

In this case there are two values of y_0 that satisfy the equation $y_0 \cosh(L/2y_0) = 1$. For example, at $L = 1$ (when the radius of the hoops equals their distance apart), Figure 25 shows that the solutions are $y_0 \simeq 0.848$ and $y_0 \simeq 0.235$. In Figure 26, we plot the corresponding solutions of equation (36), namely

$$y_c(x) \simeq 0.848 \cosh\left(\frac{x}{0.848}\right) \quad \text{and} \quad y_d(x) \simeq 0.235 \cosh\left(\frac{x}{0.235}\right)$$

(note that $y_c(0) \simeq 0.848$ and $y_d(0) \simeq 0.235$).

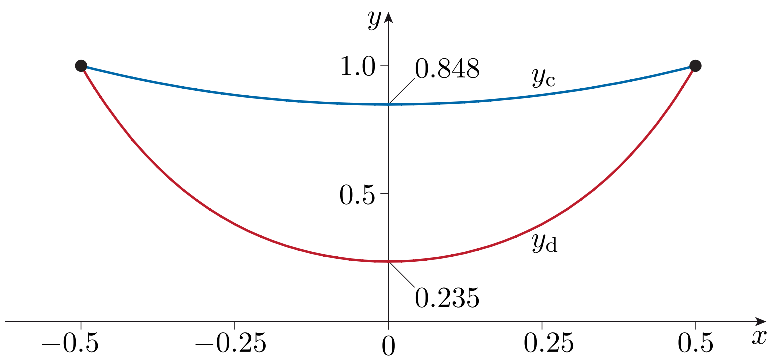


Figure 26 Graph of y_c and y_d

We see that y_{d} has a steep profile, so the shape of the soap film is like a dumbbell, while y_{c} has a shallow profile, so the soap film approximates a cylinder. These results were derived with $L = 1$, but they are typical for other values $L < 1.3255$; solutions with smaller values of y_0 (shown in red in Figure 25) correspond to dumbbell-like solutions, and those with larger values of y_0 (shown in blue in Figure 25) correspond to cylinder-like solutions. But which solutions corresponds to real soap films; that is, which solutions have minimal surface area?

The areas of the cylinder-like and dumbbell-like solutions, as functions of the distance L between the hoops, are plotted in Figure 28. You are asked to reproduce this graph in Exercise 22. From the graph, we see that the cylinder-like solutions always have smaller area than the dumbbell-like solutions, so they give the shapes of soap films. It turns out that the dumbbell-like solutions are not even local minima (they are points of inflection), but the proof of this is beyond the scope of this module.

Case 2: $L > 1.3255$

When the distance L between the hoops exceeds 1.3255, there is no solution of the equation $y_0 \cosh(L/2y_0) = 1$, so there is no smooth function $y(x)$ that gives a minimal surface of revolution. If we think of the hoops as very far apart, then there is an obvious *non*-smooth function $y(x)$ that gives rise to a surface of revolution of minimal area – the function whose graph is shown in Figure 27. This function is called the **Goldschmidt discontinuous solution**, or just the **Goldschmidt solution**, and it corresponds to a flat soap film over each hoop. The area of this solution is just the area of both hoops, namely $2\pi \simeq 6.2832$. This area is independent of the distance between the hoops, and it is represented by the green horizontal line in Figure 28. In Exercise 22 we show numerically that this line crosses the cylinder-like line at $L \simeq 1.0554$.

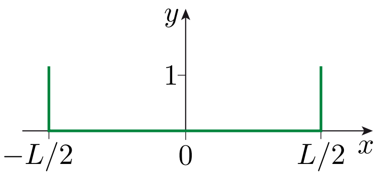


Figure 27 Graph of the Goldschmidt solution

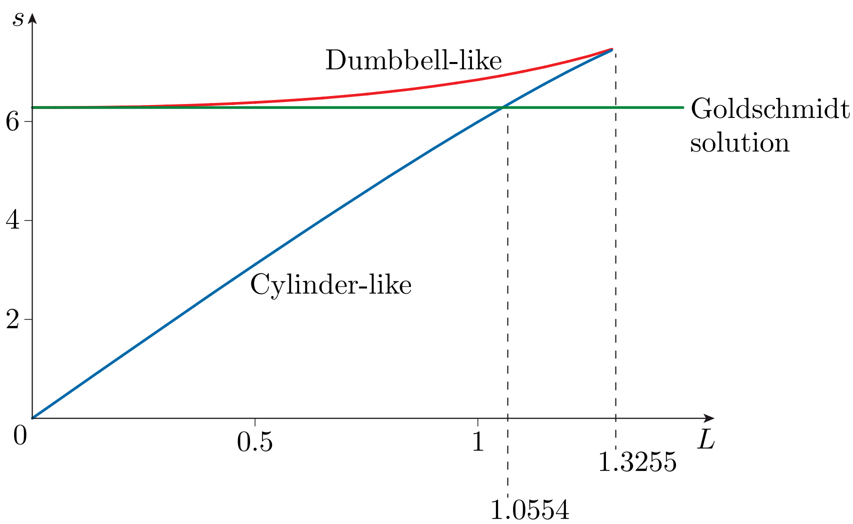


Figure 28 Graph of the surface areas of the dumbbell-like, cylinder-like and Goldschmidt solutions as functions of the distance L between the hoops

Notice that because the Goldschmidt solution is not a smooth function (it is not differentiable at $x = \pm L/2$), it cannot be obtained from the Euler–Lagrange equation.

Physical interpretation of the minimal surfaces of revolution

Let us summarise what we have learned about soap films forming minimal surfaces of revolution, with reference to Figure 28. Suppose that we dip two circular wire hoops of unit radius into a soap solution, and gradually pull the hoops apart, keeping them parallel. What does the preceding analysis predict will happen?

When the distance between the hoops is $L < 1.0554$, the soap film will assume the cylinder-like solution of equation (36), as shown in Figure 29(a).

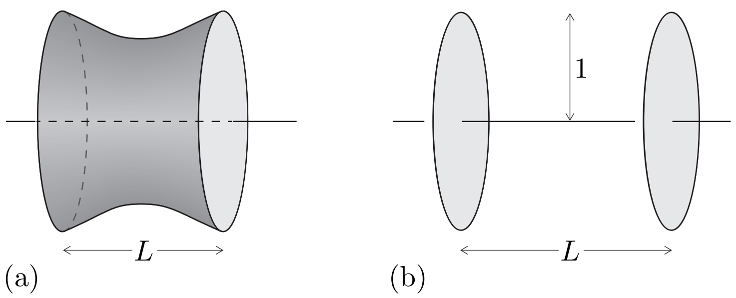


Figure 29 Soap film corresponding to: (a) the cylinder-like solution; (b) the Goldschmidt solution

As we move the rings farther apart, and once $L > 1.3255$, there will no longer be a minimal surface area that joins the two rings, and the soap film will burst – either disappearing altogether, or assuming the Goldschmidt solution shown in Figure 29(b), which is the minimal surface area solution.

For $1.0554 < L < 1.3255$, the Goldschmidt solution exists and has the smallest area. However, with these values of L, the cylinder-like solution can exist, but it is only a *local* minimum, so small perturbations in the shape of the cylinder give surfaces of larger area, but larger perturbations in the shape of the cylinder can give rise to surfaces of smaller area.

We carried out the preceding analysis for circular hoops of radius 1. But since we never specified any units in our problem, it should be clear that this analysis carries through for hoops of any radius R, provided that we scale all distances by the number R.

This relatively simple problem of finding the minimal area of a soap film provides some idea of the possible complications that can arise in variational problems.

Exercise 20

Apply the boundary conditions $y(-L/2) = 1$, $y(L/2) = 1$, to the equation

$$y = y_0 \cosh\left(\frac{x+k}{y_0}\right),$$

which is equation (35), to deduce that $k = 0$.

(*Hint*: Use the identities

$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$$

and

$$\cosh(x-y) = \cosh x \cosh y - \sinh x \sinh y.)$$

Exercise 21

As you have seen, the area of the soap film with hoops of unit radius separated by a distance L is given by

$$S[y] = 2\pi \int_{-L/2}^{L/2} y\sqrt{1+(y')^2}\,dx, \quad \text{where } y(x) = y_0 \cosh\left(\frac{x}{y_0}\right).$$

Show that

$$S[y] = \pi y_0^2 \left(\sinh\left(\frac{L}{y_0}\right) + \frac{L}{y_0}\right). \tag{37}$$

(*Hint*: Use the identities

$$\cosh^2 x - \sinh^2 x = 1 \quad \text{and} \quad \cosh^2 x = \tfrac{1}{2}(\cosh 2x + 1).)$$

Exercise 22 *Computing – optional*

(a) Use the computer software for this module to plot L against y_0, where $y_0 \cosh(L/(2y_0)) = 1$ (see Figure 25).

(b) Show numerically that this curve has a maximum $L_{\max} \simeq 1.3255$ at $y_0 \simeq 0.5524$.

(c) Read the section 'Plotting stationary curves' in the *Computer Exploration Worksheet*. Use the functions $yc0(L)$ and $yd0(L)$ defined there, and equation (37) of Exercise 21, to create a plot similar to Figure 28.

(d) Show that the area of the Goldschmidt solution and the area of the cylinder-like solution are equal when $L \simeq 1.0554$.

More about soap films

You have seen that the behaviour of soap films can be complicated, even for simple surfaces such as those formed between two hoops. The complexity of their behaviour is further compounded when one realises that there can be minimal energy solutions of a quite unexpected form, like that shown in Figure 30. We do not expect the theory described in this subsection to find such a solution because the mathematical formulation of the physical problem makes no allowance for this type of behaviour.

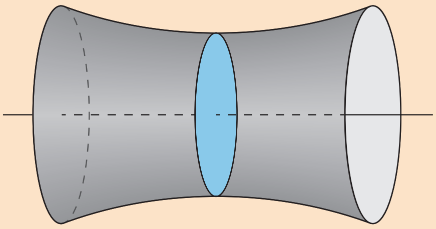


Figure 30 A physically possible soap film, produced when a circular film, perpendicular to the axis, is formed in the centre of a soap film held by two circular hoops

Soap films can form a variety of complex shapes with other boundary conditions. For example, consider a cubical frame of wire dipped into a soap solution and then taken out. Films of local minimum energy – that is, minimum area – will form on this frame. It transpires that the possible shapes formed are many and varied, and they are often counter-intuitive. Some of the possible shapes are shown in Figure 31.

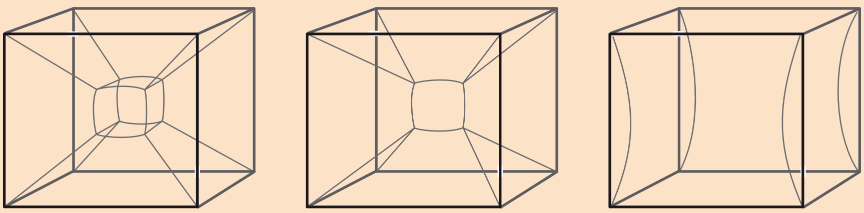


Figure 31 Soap films on cubical frames of wire

These examples illustrate the fact that there may be more than one minimal surface with the same boundary conditions. They also demonstrate that some physical problems that are simple to state can have bizarre solutions, which are difficult to describe mathematically.

We finish with a discussion of other types of functionals that are used in mathematics and physics.

Further developments in the calculus of variations

In this unit we have considered functionals of the type

$$S[y] = \int_a^b F(x, y, y')\, dx, \quad y(a) = A,\ y(b) = B, \tag{38}$$

which involve one independent variable x, a single dependent variable $y(x)$, and its first derivative $y'(x)$. There are many useful and important extensions to this type of functional, some of which we mention here.

- A simple generalisation of equation (38) is to integrands that depend on several dependent variables $y_k(x)$, $k = 1, 2, \ldots, n$, and their first derivatives, so that these have the form $F(x, y_1, y_1', y_2, y_2', \ldots, y_n, y_n')$. Functionals of this type will be studied in the next unit, on Lagrangian mechanics.
- Another generalisation involves integrands that are functions of second or higher derivatives of y, for example $F(x, y', y'')$. Such functionals are occasionally encountered in practical problems; for example, in engineering they are used to describe the mechanics of beams under stress. The differential equations that describe the stationary paths of these functionals are higher than second order. However, such functionals are not the norm, so we do not consider them any further.
- An important generalisation involves integrands that depend on two or more independent variables. For example, if there are two independent variables x_1 and x_2, then the dependent variable y is a function of both of these, and the generalisation of equation (38) has the form

  $$J[y] = \int\!\!\int_{\mathcal{D}} F\left(x_1, x_2, y, \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}\right) dx_1\, dx_2.$$

  In this case the integral defining the functional is over a region $\mathcal{D}$ in the (x_1, x_2)-plane, rather than over a line as in (38). Now the Euler–Lagrange equation becomes a partial differential equation. Such functionals and their generalisations belong to a branch of mathematical physics called **classical field theory**, which describes many important areas of physics such as the wave equation, electromagnetism and general relativity.
- One can also consider **broken extremals**, which are continuous solutions of the Euler–Lagrange equation with first derivatives that are discontinuous at a finite number of points. That such solutions are important is clear by observing collections of soap bubbles that form a composite shape with sharp corners. A simple example of such a solution is the Goldschmidt solution.

Solutions to exercises

Solution to Exercise 1

(a) On this straight line $y = 0$, so the value of the functional is

$$S[y] = \int_{-1}^{1} \sqrt{1+0}\,dx = \big[x\big]_{-1}^{1} = 1-(-1) = 2.$$

(b) As $y = \sqrt{1-x^2}$, we have

$$\frac{dy}{dx} = -\frac{x}{\sqrt{1-x^2}},$$

so the value of the functional is

$$S[y] = \int_{-1}^{1} \sqrt{1+\frac{x^2}{1-x^2}}\,dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx.$$

Using the standard integral given in the hint, we obtain

$$S[y] = \big[\arcsin x\big]_{-1}^{1} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$

Solution to Exercise 2

(a) If $y = x$, then $y' = 1$ and the functional is

$$S[y] = \int_0^1 dx = 1.$$

(b) If $y = x^3$, then $y' = 3x^2$ and the functional is

$$S[y] = 9\int_0^1 x^4\,dx = \tfrac{9}{5}.$$

Solution to Exercise 3

(a) In this case the endpoints (a,A) and (b,B) of the paths satisfy $a = 1$, $A = 0$ and $b = 2$, $B = 1$, so

$$\begin{aligned} S[y+\varepsilon g] &= \int_1^2 x((y+\varepsilon g)')^2\,dx \\ &= \int_1^2 x(y'+\varepsilon g')^2\,dx. \end{aligned}$$

To find the stationary function, we use equation (4), giving

$$\frac{d}{d\varepsilon}S[y+\varepsilon g]\bigg|_{\varepsilon=0} = \left(\frac{d}{d\varepsilon}\int_1^2 x(y'+\varepsilon g')^2\,dx\right)\bigg|_{\varepsilon=0}.$$

The integration limits 1 and 2 are independent of ε, so the order of integration and differentiation can be interchanged using Leibniz's integral rule. This gives

$$\begin{aligned}\frac{d}{d\varepsilon}S[y+\varepsilon g]\Big|_{\varepsilon=0} &= \int_1^2 \left(\frac{d}{d\varepsilon}x(y'+\varepsilon g')^2\right)\Big|_{\varepsilon=0} dx \\ &= 2\int_1^2 \left(xg'(y'+\varepsilon g')\right)\big|_{\varepsilon=0} dx \\ &= 2\int_1^2 xy'g'\,dx.\end{aligned}$$

If $S[y]$ is stationary, then it follows, by definition, that

$$\int_1^2 x\,y'(x)\,g'(x)\,dx = 0$$

for all functions $g(x)$ for which $g(1) = g(2) = 0$.

To solve this equation, we integrate by parts:

$$\int_1^2 u(x)\,g'(x)\,dx = \big[u(x)\,g(x)\big]_1^2 - \int_1^2 u'(x)\,g(x)\,dx,$$

with $u = xy'$. However, the first term on the right-hand side vanishes because $g(1) = g(2) = 0$, so we get

$$\int_1^2 (xy')'\,g(x)\,dx = 0.$$

This integral can vanish only if $(xy')' = 0$, which gives

$$x\,y'(x) = \alpha,$$

where α is a constant.

(b) The equation $xy' = \alpha$ is separable and is equivalent to

$$\frac{dy}{dx} = \frac{\alpha}{x}.$$

Integrating both sides with respect to x gives $y = \alpha \ln x + \beta$, where β is a constant. But $y(1) = 0$, so $\beta = 0$, and $y(2) = 1$, so $\alpha = 1/\ln 2$. Therefore the required solution is $y = \ln x/\ln 2$.

Solution to Exercise 4

(a) The partial derivatives are

$$\frac{\partial F}{\partial x} = \sqrt{y^2+(y')^2}, \quad \frac{\partial F}{\partial y} = \frac{xy}{\sqrt{y^2+(y')^2}}, \quad \frac{\partial F}{\partial y'} = \frac{xy'}{\sqrt{y^2+(y')^2}}.$$

The total derivative is therefore

$$\begin{aligned}\frac{dF}{dx} &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y'' \\ &= \sqrt{y^2+(y')^2} + \frac{xyy'}{\sqrt{y^2+(y')^2}} + \frac{xy'y''}{\sqrt{y^2+(y')^2}} \\ &= \frac{y^2+(y')^2+xyy'+xy'y''}{\sqrt{y^2+(y')^2}}.\end{aligned}$$

(b) (i) If $y = \sin x$, then $y' = \cos x$ and $y'' = -\sin x$.

Substituting these values into the formula for F gives $F = x\sqrt{\sin^2 x + \cos^2 x} = x$ (where we have used the identity $\sin^2 x + \cos^2 x = 1$). Hence $dF/dx = 1$.

(ii) Using the result from part (a), we obtain

$$\begin{aligned}\frac{dF}{dx} &= \frac{\sin^2 x + \cos^2 x + x\sin x\cos x - x\cos x\sin x}{\sqrt{\sin^2 x + \cos^2 x}}\\ &= \frac{1}{\sqrt{1}} = 1,\end{aligned}$$

in agreement with part (b)(i).

Solution to Exercise 5

The partial derivatives are

$$\frac{\partial F}{\partial x} = \frac{x}{\sqrt{x^2 + y(y')^2}}, \quad \frac{\partial F}{\partial y} = \frac{(y')^2}{2\sqrt{x^2 + y(y')^2}}, \quad \frac{\partial F}{\partial y'} = \frac{yy'}{\sqrt{x^2 + y(y')^2}}.$$

The total derivative is

$$\begin{aligned}\frac{dF}{dx} &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y''\\ &= \frac{x}{\sqrt{x^2 + y(y')^2}} + \frac{(y')^3}{2\sqrt{x^2 + y(y')^2}} + \frac{yy'y''}{\sqrt{x^2 + y(y')^2}}.\end{aligned}$$

Solution to Exercise 6

The partial derivatives are

$$\frac{\partial F}{\partial x} = 2xy^3 \quad \text{and} \quad \frac{\partial F}{\partial y} = 3x^2y^2,$$

and the total derivative is

$$\frac{dF}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' = 2xy^3 + 3x^2y^2y'.$$

Now, $\dfrac{d}{dx}\left(\dfrac{\partial F}{\partial y}\right)$ is the total derivative of $\dfrac{\partial F}{\partial y}$, which itself is a function of x and y. Hence

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y}\right) = \frac{\partial}{\partial x}(3x^2y^2) + y'\frac{\partial}{\partial y}(3x^2y^2) = 6xy^2 + 6x^2yy'.$$

Also, using the expression above for dF/dx, we obtain

$$\frac{\partial}{\partial y}\left(\frac{dF}{dx}\right) = 6xy^2 + 6x^2yy'.$$

Solution to Exercise 7

(a) With the notation of equation (18), the integrand is given by

$$F(x, y, y') = (y')^2 + y,$$

with $a = 0$, $b = 1$ and $A = 0$, $B = 2$. As

$$\frac{\partial F}{\partial y'} = 2y' \quad \text{and} \quad \frac{\partial F}{\partial y} = 1,$$

the Euler–Lagrange equation (19) is

$$2y'' - 1 = 0, \quad y(0) = 0,\ y(1) = 2.$$

(b) The differential equation can be written as

$$y'' = \tfrac{1}{2},$$

which can be solved by direct integration. Integrating both sides with respect to x gives $y' = \frac{1}{2}x + c$, and integrating again gives the general solution

$$y = \tfrac{1}{4}x^2 + cx + d,$$

where c and d are constants. As $y(0) = 0$, we see that $d = 0$. As $y(1) = 2$, we see that $\frac{1}{4} + c = 2$, so $c = \frac{7}{4}$. Therefore the stationary path is

$$y = \tfrac{1}{4}x^2 + \tfrac{7}{4}x.$$

Solution to Exercise 8

(a) In this case $F(x, y, y') = (y')^2 - y^2$, so

$$\frac{\partial F}{\partial y'} = 2y' \quad \text{and} \quad \frac{\partial F}{\partial y} = -2y.$$

This gives the Euler–Lagrange equation $2y'' - (-2y) = 0$, or equivalently,

$$y'' + y = 0.$$

(b) The equation $y'' + y = 0$ models a simple harmonic oscillator (see Unit 2), and has general solution

$$y = A\cos x + B\sin x,$$

where A and B are constants determined by the boundary conditions. As $y(0) = 0$, we see that $A = 0$. As $y(X) = 1$, we see that $B\sin X = 1$, so $B = 1/\sin X$. Therefore the stationary function is

$$y = \frac{\sin x}{\sin X}.$$

Solution to Exercise 9

(a) In this case $F = (y')^2 + y^2 + 2xy$, so

$$\frac{\partial F}{\partial y'} = 2y' \quad \text{and} \quad \frac{\partial F}{\partial y} = 2y + 2x.$$

This gives the Euler–Lagrange equation $2y'' - (2y + 2x) = 0$, or equivalently,

$$y'' - y = x.$$

(b) The equation $y'' - y = x$ is a second-order inhomogeneous equation, of the type that you studied in Unit 1. The solution is given by $y = y_{\text{c}} + y_{\text{p}}$, where y_{c} is the complementary function of the differential equation, and y_{p} is a particular integral.

The auxiliary equation is $\lambda^2 - 1 = 0$, with solutions $\lambda = \pm 1$, so the general solution of the homogeneous equation $y'' - y = 0$ is $y_{\text{c}} = Ce^x + De^{-x}$, where C and D are constants.

The inhomogeneous part is the expression x, so to find a particular integral we try $y_{\text{p}} = Ax + B$. Substituting this into the differential equation $y'' - y = x$ gives $-Ax - B = x$. Comparing coefficients, we see that $A = -1$ and $B = 0$, so $y_{\text{p}} = -x$.

Therefore the general solution of the equation $y'' - y = x$ is

$$y = Ce^x + De^{-x} - x.$$

To find C and D, we use the boundary conditions. As $y(0) = 0$, we see that $C + D = 0$, so $y = C(e^x - e^{-x}) - x$. As $y(1) = \alpha$, we see that $\alpha = C(e - e^{-1}) - 1$. Thus

$$y = \left(\frac{\alpha + 1}{e - e^{-1}}\right)(e^x - e^{-x}) - x.$$

If you are familiar with hyperbolic functions (a brief summary of their properties is given in Subsection 5.3), then you can simplify this expression for y. Since $\sinh x = \frac{1}{2}(e^x - e^{-x})$, we can write

$$y = (\alpha + 1)\frac{\sinh x}{\sinh 1} - x.$$

Solution to Exercise 10

(a) The integrand $F(y, y') = \ln y' + y$ does not depend on x, so it is valid to find the first integral of the Euler–Lagrange equation. As $\partial F/\partial y' = 1/y'$, we can use equation (25) to obtain

$$\frac{y'}{y'} - (\ln y' + y) = c,$$

for some constant c. Rearranging this equation, we get

$$\ln y' = 1 - c - y.$$

By taking the exponential of both sides, we obtain

$$\frac{dy}{dx} = e^{1-c-y} = e^{1-c}e^{-y} = \alpha e^{-y},$$

where $\alpha = e^{1-c}$ is another constant.

(b) The differential equation $dy/dx = \alpha e^{-y}$ is separable. Separating the variables gives

$$\int e^y\,dy = \int \alpha\,dx,$$

hence

$$e^y = \alpha x + \beta.$$

Taking the logarithm of both sides gives

$$y = \ln(\alpha x + \beta).$$

The boundary conditions tell us that

$$\ln(\alpha + \beta) = \ln 2 \quad \text{and} \quad \ln(2\alpha + \beta) = \ln 3;$$

that is, $\alpha + \beta = 2$ and $2\alpha + \beta = 3$. Therefore $\alpha = \beta = 1$. It follows that the stationary function is

$$y = \ln(x + 1).$$

Solution to Exercise 11

(a) The integrand $F(y, y') = \sqrt{y(1 + (y')^2)}$ does not depend explicitly on x, so it is valid to find the first integral of the Euler–Lagrange equation. We obtain

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

for some constant k. Simplifying this gives

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

Now square both sides to obtain

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

so

$$y' = \pm\sqrt{\frac{y}{k^2} - 1} = \frac{1}{k}\sqrt{y - k^2},$$

where in the final expression we have absorbed the factor of ± 1 into the constant k, which could be positive or negative (but not zero).

(b) The equation for y' is separable. Separating variables, we obtain

$$k\int \frac{dy}{\sqrt{y-k^2}} = \int dx.$$

Integrating both sides gives $2k\sqrt{y-k^2} = x+\alpha$, for some constant α, so

$$y = k^2 + \frac{(x+\alpha)^2}{4k^2}.$$

The boundary conditions at $x = \pm 1$ give

$$A = k^2 + \frac{(\alpha+1)^2}{4k^2} \quad \text{and} \quad A = k^2 + \frac{(\alpha-1)^2}{4k^2}.$$

By inspection (or by solving for α), these equations can be true only if $\alpha = 0$, so

$$y = k^2 + \frac{x^2}{4k^2} \quad \text{and} \quad A = k^2 + \frac{1}{4k^2}.$$

By rearranging the latter equation, we obtain $h^2 - Ah + \frac{1}{4} = 0$, where $h = k^2$, which is a quadratic equation in h that has solutions $\frac{1}{2}\left(A \pm \sqrt{A^2-1}\right)$. Therefore the stationary paths are given by

$$y(x) = h + \frac{x^2}{4h}, \quad \text{where } h = \tfrac{1}{2}\left(A \pm \sqrt{A^2-1}\right).$$

If $A > 1$, then there are two stationary paths, but if $A = 1$, then there is only one stationary path.

Solution to Exercise 12

(a) Evaluating the derivative on the left-hand side, we obtain

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

Now,

$$\frac{dF}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y'' = \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y'',$$

because $\partial F/\partial x = 0$. Combining these two equations, we find that

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

as required.

(b) The right-hand side of the preceding equation is zero if either (i) $y(x)$ satisfies the Euler–Lagrange equation

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0,$$

or (ii) $y(x)$ is constant (so that $y'(x) = 0$). In both cases, we can integrate to obtain

$$y'\frac{\partial F}{\partial y'} - F = c,$$

for some constant c. This is the first-order differential equation of equation (25).

Solution to Exercise 13

The derivative of $f(x)$ is

$$f'(x) = \frac{x}{\sqrt{x^2 + h_1^2}} - \frac{d - x}{\sqrt{(d - x)^2 + h_2^2}}.$$

However, from Figure 12 we have

$$\sin\theta_1 = \frac{x}{\sqrt{x^2 + h_1^2}} \quad \text{and} \quad \sin\theta_2 = \frac{d - x}{\sqrt{(d - x)^2 + h_2^2}}.$$

It follows that

$$f'(x) = \sin\theta_1 - \sin\theta_2.$$

Therefore the function $f(x)$ is stationary when $\sin\theta_1 = \sin\theta_2$, that is, $\theta_1 = \theta_2$. Furthermore, as

$$f''(x) = \frac{h_1^2}{(x^2 + h_1^2)^{3/2}} + \frac{h_2^2}{((d - x)^2 + h_2^2)^{3/2}} > 0,$$

the stationary point is a minimum.

Solution to Exercise 14

(a) From Fermat's principle, the light ray will be a stationary path of the functional

$$T[y] = \frac{2}{c}\int_a^b \frac{1}{y}\sqrt{1 + (y')^2}\,dx, \quad y(a) = A,\ y(b) = B.$$

As the integrand

$$F = \frac{2}{cy}\sqrt{1 + (y')^2}$$

does not depend on the independent variable x, we can use the first integral of the Euler–Lagrange equation (see equation (25)) to obtain

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

where k is a constant into which we have absorbed the common factor $2/c$. Multiplying throughout by $y\sqrt{1 + (y')^2}$ and simplifying gives $ky\sqrt{1 + (y')^2} = -1$. Then square both sides and solve for y' to obtain

$$\begin{aligned}\frac{dy}{dx} &= \pm\sqrt{\frac{1}{k^2y^2} - 1}\\ &= \pm\frac{1}{ky}\sqrt{1 - k^2y^2}.\end{aligned}$$

Because k is an arbitrary constant, which could be positive or negative (but not zero), we can remove the factor ±1 to obtain

$$\frac{dy}{dx} = \frac{1}{ky}\sqrt{1 - k^2y^2}, \quad \text{where } \tfrac{1}{2} \leq y \leq 2.$$

(b) The preceding differential equation is separable. Separating variables gives

$$\int \frac{ky}{\sqrt{1-k^2y^2}}\,dy = \int dx.$$

As

$$\frac{d}{dy}\sqrt{1-k^2y^2} = -\frac{k^2y}{\sqrt{1-k^2y^2}},$$

we can integrate both sides of the equation above to give

$$-\frac{1}{k}\sqrt{1-k^2y^2} = x+\alpha,$$

where α is a constant. Squaring both sides gives $1/k^2 - y^2 = (x+\alpha)^2$, thus

$$y = \sqrt{\frac{1}{k^2}-(x+\alpha)^2},$$

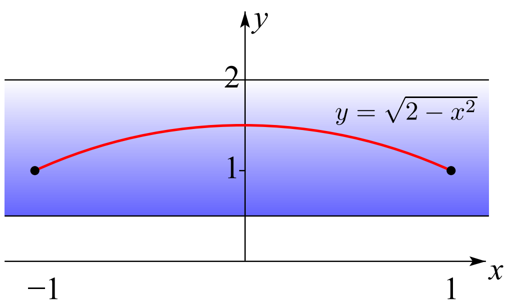
where we take the positive sign because $y \geq \frac{1}{2}$.

To find k and α, we apply the boundary conditions $y(-1) = y(1) = 1$. We obtain

$$\sqrt{\frac{1}{k^2}-(\alpha-1)^2} = 1 \quad \text{and} \quad \sqrt{\frac{1}{k^2}-(\alpha+1)^2} = 1.$$

By inspection, these equations can both be true only if $\alpha = 0$ (you can also show this by squaring both sides of these equations and solving for α). So $\sqrt{1/k^2-1} = 1$. Solving this equation for k gives $k = \pm 1/\sqrt{2}$, and hence the path of the light ray (as shown in the figure) is

$$y = \sqrt{2-x^2}, \quad \text{where } -1 \leq x \leq 1.$$



Solution to Exercise 15

As $y = x/b$, we have $y' = 1/b$, so

$$\sqrt{\frac{1+(y')^2}{y}} = \sqrt{\frac{1+1/b^2}{x/b}} = \sqrt{\frac{b^2+1}{b}}\,\frac{1}{\sqrt{x}}.$$

Therefore equation (29) tells us that the time T satisfies

$$T = \sqrt{\frac{b^2+1}{2gb}} \int_0^b \frac{1}{\sqrt{x}}\,dx = 2\sqrt{\frac{b^2+1}{2gb}}\,\left[\sqrt{x}\,\right]_0^b = 2\sqrt{\frac{b^2+1}{2g}}.$$

Solution to Exercise 16

(a) Equation (31) says that

$$x = \frac{2\phi - \sin 2\phi}{1-\cos 2\phi_b} \quad \text{and} \quad y = \frac{1-\cos 2\phi}{1-\cos 2\phi_b}.$$

Therefore the gradient of the curve $y = y(x)$ is

$$\frac{dy}{dx} = \frac{dy}{d\phi}\bigg/\frac{dx}{d\phi} = \frac{2\sin 2\phi}{2(1-\cos 2\phi)} = \frac{2\sin\phi\cos\phi}{2\sin^2\phi} = \frac{1}{\tan\phi}.$$

(b) At the origin, we have $x = y = 0$, so $\phi = 0$. Therefore $\tan\phi = 0$, which implies that dy/dx is infinite, so the brachistochrone hangs vertically downwards at $(0,0)$.

(c) The brachistochrone is parallel to the x-axis at points where the gradient is 0, that is, when $\tan\phi$ is infinite. This occurs only when $\phi = \pi/2$, in our range. At this value of ϕ, we have

$$x = \frac{\pi}{1-\cos 2\phi_b} \quad \text{and} \quad y = \frac{2}{1-\cos 2\phi_b}.$$

The brachistochrone terminates at $(x,y) = (b,1)$, so the critical brachistochrone must satisfy

$$\frac{\pi}{1-\cos 2\phi_b} = b \quad \text{and} \quad \frac{2}{1-\cos 2\phi_b} = 1.$$

Dividing one equation by the other gives $b = \pi/2$, as required.

Solution to Exercise 17

(a) In this example the integrand of the functional has the form $F(x,y') = \sqrt{1+x+(y')^2}$, and

$$\frac{\partial F}{\partial y} = 0 \quad \text{and} \quad \frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{1+x+(y')^2}}.$$

Hence the Euler–Lagrange equation (19) becomes

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

Integrating both sides with respect to x gives

$$y' = c\sqrt{1+x+(y')^2},$$

for some constant c, as required.

(b) Squaring and rearranging the preceding equation gives

$$\left(\frac{dy}{dx}\right)^2 = k^2(1+x), \quad \text{where } k^2 = \frac{c^2}{1-c^2}.$$

Taking square roots, we obtain $y' = \pm k(1+x)$, but we may as well write this as

$$y' = k(1+x),$$

because k is a constant that could be positive or negative. By separating variables in this equation, we obtain

$$\int dy = k \int \sqrt{1+x}\, dx,$$

so

$$y = \alpha(1+x)^{3/2} + \beta,$$

where $\alpha = \frac{2}{3}k$ and β are constants. Applying the boundary conditions gives

$$\alpha + \beta = 0 \quad \text{and} \quad 2^{3/2}\alpha + \beta = 1,$$

which have solution $\alpha = -\beta = 1/(2^{3/2} - 1)$. Hence

$$y = \frac{(1+x)^{3/2} - 1}{2^{3/2} - 1}.$$

Solution to Exercise 19

(a) The general equation of a non-vertical straight line is $y = mx + c$, where m and c are constants. The line passes through $(0,0)$, giving $c = 0$, and through (b, B), giving $m = B/b$. Therefore

$$y = \frac{B}{b}x.$$

(b) Substituting the preceding equation into equation (33) gives

$$\begin{aligned} S[y] &= 2\pi \int_0^b \frac{B}{b}x\sqrt{1 + \frac{B^2}{b^2}}\, dx \\ &= \pi \frac{B}{b^2}\sqrt{b^2 + B^2}\,\big[x^2\big]_0^b \\ &= \pi B\sqrt{b^2 + B^2}. \end{aligned}$$

Solution to Exercise 20

The boundary conditions give

$$y_0 \cosh\left(\frac{2k+L}{2y_0}\right) = 1 \quad \text{and} \quad y_0 \cosh\left(\frac{2k-L}{2y_0}\right) = 1.$$

Because cosh is an even function, it should be fairly clear that these equations can be satisfied only if $k = 0$. However, let us establish this fact algebraically.

Using the identity given in the hint with $x = k/y_0$ and $y = L/(2y_0)$, we obtain

$$\begin{aligned} &y_0 \cosh(k/y_0)\cosh(L/(2y_0)) + y_0 \sinh(k/y_0)\sinh(L/(2y_0)) = 1, \\ &y_0 \cosh(k/y_0)\cosh(L/(2y_0)) - y_0 \sinh(k/y_0)\sinh(L/(2y_0)) = 1. \end{aligned}$$

Subtracting the second equation from the first gives $2y_0 \sinh(k/y_0)\sinh(L/2y_0) = 0$, which, because $L > 0$, can be true only if $y_0 \sinh(k/y_0) = 0$. As y_0 is not zero, we see that $\sinh(k/y_0) = 0$, so $k = 0$.

Solution to Exercise 21

Observe that $y'(x) = \sinh(x/y_0)$, so

$$\begin{aligned} S[y] &= 2\pi y_0 \int_{-L/2}^{L/2} \cosh\left(\frac{x}{y_0}\right)\sqrt{1+\sinh^2\left(\frac{x}{y_0}\right)}\,dx \\ &= 2\pi y_0 \int_{-L/2}^{L/2} \cosh^2\left(\frac{x}{y_0}\right) dx, \end{aligned}$$

where we have used the first identity from the hint. Now use the second identity from the hint to obtain

$$\begin{aligned} S[y] &= \pi y_0 \int_{-L/2}^{L/2} \left(\cosh\left(\frac{2x}{y_0}\right) + 1\right) dx \\ &= \pi y_0 \left[\frac{y_0}{2}\sinh\left(\frac{2x}{y_0}\right) + x\right]_{-L/2}^{L/2} \\ &= \pi y_0^2 \left(\sinh\left(\frac{L}{y_0}\right) + \frac{L}{y_0}\right), \end{aligned}$$

where, to determine the last line, we have used the fact that sinh is an odd function.

Acknowledgements

Grateful acknowledgement is made to the following sources:

Figure 9: © Soapbubble.dk. This file is licensed under the Creative Commons Attribution-Noncommercial Licence http://creativecommons.org/licenses/by-nc/3.0.

Figure 13: Brocken Inaglory. This file is licensed under the Creative Commons Attribution-Share Alike Licence http://creativecommons.org/licenses/by-sa/3.0.

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.

Index