Chapter 2

The Calculus of Variations

label: chap:vp2

2.1 Introduction

In this chapter we consider the particular variational principle defining the shortest distance between two points in a plane. It is well known that this shortest path 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 essential ideas of the Calculus of Variations by finding the equation of this line. The algebra may seem overcomplicated for this simple problem, but the same theory can be applied to far more complicated problems, and we shall see in chapter 3 the most important equation of the Calculus of Variations, the Euler-Lagrange equation, can be derived with almost no extra effort.

The chapter ends with a description of some of the problems that can be formulated in terms of variational principles, some of which will be solved later in the course.

The approach adopted is intuitive, that is we assume that functionals behave like functions of n real variables. This is exactly the approach used by Euler (1707-1783) and Lagrange (1736-1813) in their original analysis and it can be successfully applied to many important problems. However, it masks a number of problems, all to do with the subtle difference between infinite and finite dimensional spaces: some of these problems will be discussed later in the course. Initially, however, they are ignored in order to progress.

2.2 The shortest distance between two points in a plane

label: sec:vp2-short

The distance between two points $P_a = (a, A)$ and $P_b = (b, B)$ in the Oxy-plane along a given curve, defined by the function $y(x)$, is given by the functional

label: eq:vp2-01

$$S[y] = \int_a^b dx \, \sqrt{1 + y'(x)^2}. \tag{2.1}$$

The curve must pass through the end points, so $y(x)$ satisfies the boundary conditions, $y(a) = A$ and $y(b) = B$. We shall usually assume that $y'(x)$ is continuous on (a, b).

We require the equation of the function that makes $S[y]$ stationary, that is we need to understand how the values of the functional $S[y]$ change as the path between P_a and P_b varies. These ideas are introduced here, and developed in chapter 3, using analogies with the theory of functions of many real variables.

2.2.1 The stationary distance

label: sec:vp2-stat

In the theory of functions of several real variables a stationary point is *defined* to be one at which the values of the function at all neighbouring points are 'almost' the same as at the stationary point. To be precise, if $G(\mathbf{x})$ is a function of n real variables, $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, we compare values of G at $\mathbf{x}$ and the nearby point $\mathbf{x} + \epsilon\boldsymbol{\xi}$, where $|\epsilon| \ll 1$ and $|\boldsymbol{\xi}| = 1$. Taylor's expansion gives,

label: eq:vp2-03a

$$G(\mathbf{x} + \epsilon\boldsymbol{\xi}) - G(\mathbf{x}) = \epsilon \sum_{k=1}^{n} \frac{\partial G}{\partial x_k}\xi_k + O(\epsilon^2). \tag{2.2}$$

A stationary point is *defined* to be one for which the term $O(\epsilon)$ is zero for *all* $\boldsymbol{\xi}$. This gives the familiar conditions for a point to be stationary, namely $\partial G/\partial x_k = 0$ for $k = 1, 2, \cdots, n$.

For a functional we proceed in the same way. That is, we choose adjacent paths joining P_a to P_b and compare the values of S along these paths. If a path is represented by a differentiable function $y(x)$, adjacent paths may be represented by $y(x) + \epsilon h(x)$, where ϵ is a real variable and $h(x)$ another differentiable function. Since all paths must pass through P_a and P_b, we require $y(a) = A$, $y(b) = B$ and $h(a) = h(b) = 0$; otherwise $h(x)$ is arbitrary. The difference

$$\delta S = S[y + \epsilon h] - S[y],$$

may be considered as a function of the real variable ϵ, for arbitrary $y(x)$ and $h(x)$ and for small values of ϵ, $|\epsilon| \ll 1$. When $\epsilon = 0$, $\delta S = 0$ and for small $|\epsilon|$ we expect δS to be proportional to ϵ; in general this is true as seen in equation 2.3 below.

However, there may be some paths for which δS is proportional to ϵ^2, rather than ϵ. These paths are special and we *define* these to be the *stationary paths*, *curves* or *stationary functions*. Thus a *necessary* condition for a path $y(x)$ to be a stationary path is that

$$S[y + \epsilon h] - S[y] = O(\epsilon^2),$$

for *all* suitable $h(x)$. The equation for the stationary function $y(x)$ is obtained by examining this difference more carefully.

The distances along these adjacent curves are

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

We proceed by expanding the integrand of $S[y + \epsilon h]$ in powers of ϵ, retaining only the terms proportional to ϵ. One way of making this expansion is to consider the integrand as a function of ϵ and to use Taylor's series to expand in powers of ϵ,

$$\begin{aligned}\sqrt{1+(y'+\epsilon h')^2} &= \sqrt{1+y'^2}+\epsilon\left[\frac{d}{d\epsilon}\sqrt{1+(y'+\epsilon h')^2}\right]_{\epsilon=0}+O(\epsilon^2),\\ &= \sqrt{1+y'^2}+\epsilon\frac{y'h'}{\sqrt{1+y'^2}}+O(\epsilon^2).\end{aligned}$$

Substituting this expansion into the integral and rearranging gives the difference between the two lengths, label: eq:vp2-04

$$S[y+\epsilon h]-S[y]=\epsilon\int_a^b dx\,\frac{y'(x)}{\sqrt{1+y'(x)^2}}h'(x)+O(\epsilon^2). \tag{2.3}$$

This difference depends upon both $y(x)$ and $h(x)$, just as for functions of n real variables the difference $G(\mathbf{x}+\epsilon\boldsymbol{\xi})-G(\mathbf{x})$, equation 2.2, depends upon both $\mathbf{x}$ and $\boldsymbol{\xi}$, the equivalents of $y(x)$ and $h(x)$ respectively.

Since $S[y]$ is stationary it follows, by definition, that label: eq:vp2-05

$$\int_a^b dx\,\frac{y'(x)}{\sqrt{1+y'(x)^2}}h'(x)=0 \tag{2.4}$$

for all suitable functions $h(x)$.

We shall see in chapter 3 that because 2.4 holds for *all* those functions $h(x)$ for which $h(a)=h(b)=0$ and $h'(x)$ is continuous, this equation is sufficient to determine $y(x)$ uniquely. Here, however, we simply show that if label: eq:vp2-06

$$\frac{y'(x)}{\sqrt{1+y'(x)^2}}=\alpha=\text{constant}\quad\text{for all}\quad x, \tag{2.5}$$

then the integral in equation 2.4 is zero for all $h(x)$. Assuming that 2.5 is true, equation 2.4 becomes

$$\int_a^b dx\,\alpha h'(x)=\alpha\left\{h(b)-h(a)\right\}=0\quad\text{since}\quad h(a)=h(b)=0.$$

In section 3.3 we show that condition 2.5 is *necessary* as well as sufficient for equation 2.4 to hold.

Equation 2.5 shows that $y'(x)=m$, where m is a constant, and integration gives the general solution,

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

for another constant c: this is the equation of a straight line as expected. The constants m and c are determined by the conditions that the straight line passes through P_a and P_b: label: eq:vp2-07

$$y(x)=\frac{B-A}{b-a}x+\frac{Ab-Ba}{b-a}. \tag{2.6}$$

This analysis shows that the functional $S[y]$ defined in equation 2.1 is *stationary* along the straight line joining P_a to P_b. We have *not* shown that this gives a minimum distance: this is proved in exercise 2.2. label: ex:vp2-02

Exercise 2.1
Use the above method on the functional

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

to show that the stationary function is the straight line $y(x) = Bx$, and that the value of the functional on this line is $S[y] = \sqrt{1 + B}$.

2.2.2 The shortest path: local and global minima

label:
sec:vp2-local

In this section we show that the straight line 2.6 gives the minimum distance. For practical reasons this analysis is divided into two stages. First, we show that the straight line is a *local* minimum of the functional, using an analysis that is generalised in chapter 6 to functionals. Second, we show that, amongst the class of differentiable functions, the straight line is actually a *global* minimum: this analysis makes use of special features of the integrand.

The distinction between local and global extrema is illustrated in figure 2.1. Here we show a function $f(x)$, defined in the interval $a \le x \le b$, having three stationary points B, C and D, two of which are minima the other being a maximum. It is clear from the figure that at the stationary point D, $f(x)$ takes its smallest value in the interval — so this is the global minimum. The function is largest at A, but this point is not stationary — this is the global maximum. The stationary points at B is a local minimum, because here, $f(x)$ is smaller than at any point in the neighbourhood of B: likewise the points C and D are local maxima and minima, respectively. The adjective local is frequently omitted. In some texts local extrema are named *relative extrema*.

label:
f:vp2-local

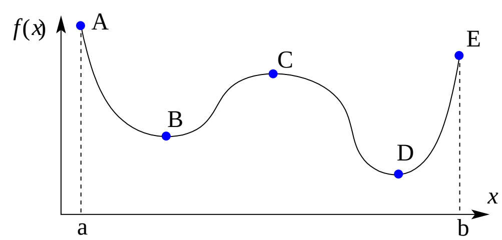

Figure 2.1 Diagram to illustrate the difference between local and global extrema.

It is clear from this example that to classify a point as a local extrema requires an examination of the function values only in the neighbourhood of the point. Whereas, determining whether a point is a global extrema requires examining all values of the function; this type of analysis usually invokes special features of the function.

The local analysis of a stationary point of a function, $G(\mathbf{x})$, of n variables proceeds by making a second-order Taylor expansion about a point $\mathbf{x} = \mathbf{a}$,

$$G(\mathbf{a} + \epsilon\boldsymbol{\xi}) = G(\mathbf{a}) + \epsilon \sum_{k=1}^{n} \frac{\partial G}{\partial x_k}\xi_k + \frac{1}{2}\epsilon^2 \sum_{k=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 G}{\partial x_k \partial x_j}\xi_k\xi_j + \cdots,$$

where all derivatives are evaluated at $\mathbf{x} = \mathbf{a}$. If $G(\mathbf{x})$ is stationary at $\mathbf{x} = \mathbf{a}$ then all first derivatives are zero. The nature of the stationary point is usually determined by the behaviour of the second-order term. For a stationary point to be a local minimum it is necessary for the quadratic terms to be strictly positive for all $\boldsymbol{\xi}$, that is

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial^2 G}{\partial x_k \partial x_j}\xi_k\xi_j > 0 \quad \text{for all} \quad \xi_k, \xi_j, \quad k, j = 1, 2, \cdots, n,$$

with $|\boldsymbol{\xi}| = 1$. The stationary point is a local maximum if this quadratic form is strictly negative. For large n it is usually difficult to determine whether these inequalities are satisfied, although there are well defined tests which are described in chapter 6.

For a functional we proceed in the same way: the nature of a stationary path is usually determined by the second-order expansion. If $S[y]$ is stationary then, by definition,

$$S[y + \epsilon h] - S[y] = \frac{1}{2}\Delta_2[y, h]\epsilon^2 + O(\epsilon^3)$$

for some quantity $\Delta_2[y, h]$, depending upon both y and h; special cases of this expansion are found in exercise 2.2 and 2.3. Then $S[y]$ is a local minimum if $\Delta_2[y, h] > 0$ for all $h(x)$, and a local maximum if $\Delta_2[y, h] < 0$ for all $h(x)$. Normally it is difficult to establish these inequalities, and the general theory is described in chapter 6. For the functional defined by equation 2.1, however, the proof is straight-forward; the following exercise guides you through it. label: ex:vp2-03

Exercise 2.2

(a) Use Taylor's series 1.26 (page 28) to obtain the following expansion in ϵ,

$$\sqrt{1 + (\alpha + \epsilon\beta)^2} = \sqrt{1 + \alpha^2} + \frac{\alpha\beta\epsilon}{\sqrt{1 + \alpha^2}} + \frac{\beta^2\epsilon^2}{2(1 + \alpha^2)^{3/2}} + O(\epsilon^3).$$

(b) Use this result to show that if $y(x)$ is the straight line defined in equation 2.6 and $S[y]$ the functional 2.1, then

$$S[y + \epsilon h] - S[y] = \frac{\epsilon^2}{2(1 + m^2)^{3/2}}\int_a^b dx\, h'(x)^2, \quad m = \frac{B - A}{b - a}.$$

Deduce that the straight line is a local minimum for the distance between P_a and P_b.

label: ex:vp2-04

Exercise 2.3

In this exercise the functional defined in exercise 2.1 is considered in more detail.

By expanding the integrand of $S[y + \epsilon h]$ to second order in ϵ show that, if $y(x)$ is the stationary path, then

$$S[y + \epsilon h] = S[y] - \frac{\epsilon^2}{8(1 + B)^{3/2}}\int_0^1 dx\, h'(x)^2, \quad B > -1.$$

Deduce that the path $y(x) = Bx$, $B > -1$, is a local maximum of this functional.

Now we show that the straight line is also a *global* minimum. This analysis relies on a special property of the integrand, which will also be useful later on in chapter 6. This property is a consequence of the Cauchy-Schwarz inequality, given on page 37. label: ex:vp2-04a

Exercise 2.4
Use the Cauchy-Schwarz inequality (page 37) with $\mathbf{a} = (1, z)$ and $\mathbf{b} = (1, z + u)$ to show that

$$\sqrt{1 + (z + u)^2}\sqrt{1 + z^2} \geq 1 + z^2 + zu$$

with equality only if $u = 0$. Hence show that

$$\sqrt{1 + (z + u)^2} - \sqrt{1 + z^2} \geq \frac{zu}{\sqrt{1 + z^2}}.$$

Using this inequality with $z = y'(x)$ and $u = \epsilon h'(x)$ we see that

$$S[y + \epsilon h] - S[y] \geq \epsilon \int_0^1 dx \, \frac{y'}{\sqrt{1 + y'^2}} \, h'.$$

If $y(x)$ is the stationary path, $y = x$, then since $y' = 1$ and $h(0) = h(1) = 0$ we have $S[y + \epsilon h] \geq S[y]$ for all nonzero $h(x)$.

This analysis did not assume that $|\epsilon|$ is small, and since all admissible paths can be expressed in the form $x + \epsilon h(x)$, we have shown that the straight line is the global minimum, for the class of differentiable functions.

An observation

Problems involving shortest distances on surfaces other than a plane illustrate other features of variational problems. Thus if we replace the plane by the surface of a sphere then the shortest distance between two points on the surface is the arc length of a great circle joining the two points — that is the circle created by the intersection of the spherical surface and the plane passing through the two points and the centre of the sphere; this problem is examined in exercise 4.20 (page 173). Now, for most points, there are two stationary paths corresponding to the long and the short arcs of the great circle. However, if the points are at opposite ends of a diameter, there are infinitely many shortest paths. This example shows that solutions to variational problems may be complicated.

In general, the stationary paths between two points on a surface are named geodesics. For a plane surface the only geodesics are straight lines; for a sphere, most pairs of points are joined by just two geodesics that are the segments of the great circle through the points. For other surfaces there may be several stationary paths: an example of the consequences of such complications is described next.

2.2.3 Gravitational Lensing

The general theory of relativity, discovered by Einstein (1879-1955), shows that the path taken by light from a source to an observer is along a geodesic on a surface in a four-dimensional space. In this theory gravitational forces are represented by distortions to this surface. The theory therefore predicts that light is "bent" by gravitational forces, a prediction that was first observed in 1919 by Eddington (1882-1944) in his measurements of the position of stars during a total solar eclipse: these observations provided the first direct confirmation of Einstein's general theory of relativity.

The departure from a straight line path depends upon the mass of the body between the source and observer. If it is sufficiently massive, two images may be seen as illustrated schematically in figure 2.2.

label: f:vp2-glense

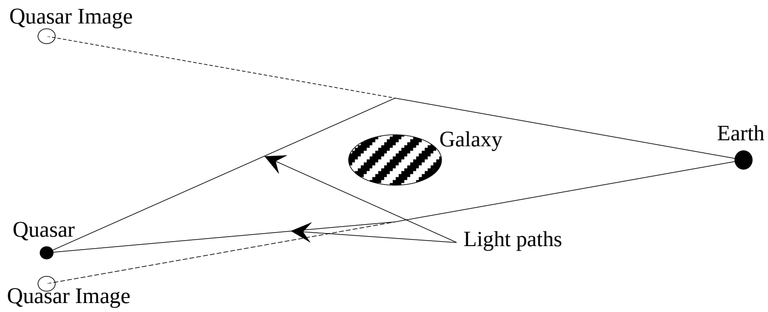

Figure 2.2 Diagram showing how an intervening galaxy can sufficiently distort a path of light from a bright object, such as a quasar, to provide two stationary paths and hence two images. Many examples of such multiple images, and more complicated but similar optical effects, have now been observed. Usually there are more than two stationary paths.

2.3 Two generalisations

2.3.1 Functionals depending only upon $y'(x)$

The functional 2.1 (page 73) depends only upon the derivative of the unknown function. Although this is a special case it is worth considering in more detail in order to develop the notation we need. label: sec:vp2-131

If $F(z)$ is a differentiable function of z then a general functional of the form of 2.1 is

label: eq:vp2-10

$$S[y] = \int_a^b dx\, F(y'), \quad y(a) = A, \quad y(b) = B, \tag{2.7}$$

where $F(y')$ simply means that in $F(z)$ all occurrences of z are replaced by $y'(x)$. Thus in the previous example $F(z) = \sqrt{1+z^2}$ so $F(y') = \sqrt{1+y'(x)^2}$. Note that the symbols $F(y')$ and $F(y'(x))$ denote the same function.

The difference between the functional evaluated along $y(x)$ and the adjacent paths $y(x) + \epsilon h(x)$, where $|\epsilon| \ll 1$ and $h(a) = h(b) = 0$, is label: eq:vp2-11

$$S[y + \epsilon h] - S[y] = \int_a^b dx\, \left\{ F(y' + \epsilon h') - F(y') \right\}. \tag{2.8}$$

Now we need to express $F(y' + \epsilon h')$ as a series in ϵ; assuming that $F(z)$ is differentiable, Taylor's theorem gives

$$F(z + \epsilon u) = F(z) + \epsilon u \frac{dF}{dz} + O(\epsilon^2).$$

The expansion of $F(y' + \epsilon h')$ is obtained from this simply by the replacements $z \to y'(x)$ and $u \to h'(x)$, which gives label: eq:vp2-12

$$F(y' + \epsilon h') - F(y') = \epsilon h'(x) \frac{d}{dy'} F(y') + O(\epsilon^2) \tag{2.9}$$

where the notation dF/dy' means label: eq:vp2-12a

$$\frac{d}{dy'}F(y') = \left.\frac{dF}{dz}\right|_{z=y'(x)}. \tag{2.10}$$

For instance, if $F(z) = \sqrt{1+z^2}$ then

$$\frac{dF}{dz} = \frac{z}{\sqrt{1+z^2}} \quad \text{and} \quad \frac{dF}{dy'} = \frac{y'(x)}{\sqrt{1+y'(x)^2}}.$$

label: ex:vp2-05

Exercise 2.5
Find the expressions for dF/dy' when

(a) $F(y') = (1+y'^2)^{1/4}$, (b) $F(y') = \sin y'$, (c) $F(y') = \exp(y')$.

label: eq:vp2-13

Substituting the difference 2.9 into the equation 2.8 gives

$$S[y+\epsilon h] - S[y] = \epsilon \int_a^b dx\, h'(x) \frac{d}{dy'}F(y') + O(\epsilon^2). \tag{2.11}$$

The functional $S[y]$ is stationary if the term $O(\epsilon)$ is zero for all suitable functions $h(x)$. As before we give a sufficient condition, deferring the proof that it is also necessary. In this analysis it is important to remember that $F(z)$ is a given function and that $y(x)$ is an unkown function that we need to find. Observe that if

label: eq:vp2-14

$$\frac{d}{dy'}F(y') = \alpha = \text{constant} \tag{2.12}$$

then

$$S[y+\epsilon h] - S[y] = \epsilon\alpha\left(h(b) - h(a)\right) + O(\epsilon^2) = O(\epsilon^2) \quad \text{since} \quad h(a) = h(b) = 0.$$

In general equation 2.12 is true only if $y'(x)$ is also constant, and hence

$$y(x) = mx + c \quad \text{and therefore} \quad y(x) = \frac{B-A}{b-a}x + \frac{Ab-Ba}{b-a},$$

the last result following from the boundary conditions $y(a) = A$ and $y(b) = B$.

This is the same solution as given in equation 2.6. Thus, for this class of functional, the stationary function is independent of the form of the integrand although its nature is not, see for instance exercise 2.17 (page 96).

The exceptional example is when $F(z)$ is linear, in which case the value of $S[y]$ depends only upon the end points and not the values of $y(x)$ in between, as shown in the following exercise.

label: ex:vp2-05a

Exercise 2.6
If $F(z) = Cz + D$, where C and D are constants, by showing that the value of the functional $S[y] = \int_a^b dx\, F(y')$ is independent of the chosen path, deduce that equation 2.12 does *not* imply that $y'(x) = $constant.

What is the effect of making either, or both C and D a function of x?

2.3.2 Functionals depending upon x and $y'(x)$

label: eq:vp2-20

Now consider the slightly more general functional

$$S[y] = \int_a^b dx\, F(x, y'), \quad y(a) = A, \quad y(b) = B, \tag{2.13}$$

where the integrand $F(x, y')$ depends explicitly upon the two variables x and y'. The difference in the value of the functional along adjacent paths is

label: eq:vp2-21

$$S[y + \epsilon h] - S[y] = \int_a^b dx\, \left\{ F(x, y' + \epsilon h') - F(x, y') \right\}. \tag{2.14}$$

In this example $F(x, z)$ is a function of two variables and we require the expansion

$$F(x, z + \epsilon u) = F(x, z) + \epsilon u \frac{\partial F}{\partial z} + O(\epsilon^2)$$

where Taylor's series for functions of two variables is used. Comparing this with the expression in equation 2.9 we see that the only difference is that the derivative with respect to y' has been replaced by a partial derivative. As before, replacing z by $y'(x)$ and u by $h'(x)$, equation 2.14 becomes

label: eq:vp2-22

$$S[y + \epsilon h] - S[y] = \epsilon \int_a^b dx\, h'(x)\, \frac{\partial}{\partial y'} F(x, y') + O(\epsilon^2). \tag{2.15}$$

If $y(x)$ is the stationary path it is necessary that

$$\int_a^b dx\, h'(x) \frac{\partial}{\partial y'} F(x, y') = 0 \quad \text{for all} \quad h(x).$$

As before a sufficient condition for this is that $F_{y'}(x, y') =$ constant, which gives the following differential equation for $y(x)$,

label: eq:vp2-23

$$\frac{\partial}{\partial y'} F(x, y') = c, \quad y(a) = A, \quad y(b) = B, \tag{2.16}$$

where c is a constant. This is the equivalent of equation 2.12, but now the explicit presence of x in the equation means that $y'(x) =$constant is *not* a solution.

label: ex:vp2-06

Exercise 2.7
Consider the functional

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

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

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

where c is a constant, makes $S[y]$ stationary. By expressing $y'(x)$ in terms of x solve this equation to show that

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

2.4 Notation

label: sec:vp2-nota

In the previous sections we used the notation $F(y')$ to denote a function of the derivative of $y(x)$ and proceeded to treat y' as an independent variable, so that the expression dF/dy' had the meaning defined in equation 2.10. This notation and its generalisation are very important in subsequent analysis; it is therefore essential that you are familiar with it and can use it.

Consider a function $F(x,u,v)$ of three variables, for instance $F = x\sqrt{u^2+v^2}$, and assume that all necessary partial derivatives of $F(x,u,v)$ exist. If $y(x)$ is a function of x we may form a function of x with the substitutions $u \to y(x)$, $v \to y'(x)$, thus

$$F(x,u,v) \quad \text{becomes} \quad F(x,y,y').$$

Depending upon circumstances $F(x,y,y')$ can be considered either as a function of a single variable x, as when evaluating the integral $\int_a^b dx\, F(x,y(x),y'(x))$, or as a function of three independent variables (x,y,y'). In the latter case the first partial derivatives with respect to y and y' are just

$$\frac{\partial F}{\partial y} = \left.\frac{\partial F}{\partial u}\right|_{u=y,v=y'} \quad \text{and} \quad \frac{\partial F}{\partial y'} = \left.\frac{\partial F}{\partial v}\right|_{u=y,v=y'}.$$

Because y depends upon x we may also form the total derivative of $F(x,y,y')$ with respect to x using the chain rule, equation 1.21 (page 24)

label: eq:vp2-not01

$$\frac{dF}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y'(x) + \frac{\partial F}{\partial y'}y''(x). \tag{2.17}$$

In the particular case $F(x,u,v) = x\sqrt{u^2+v^2}$ these rules give

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

Similarly, the second-order derivatives are

$$\frac{\partial^2 F}{\partial y^2} = \left.\frac{\partial^2 F}{\partial u^2}\right|_{u=y,v=y'}, \quad \frac{\partial^2 F}{\partial y'^2} = \left.\frac{\partial^2 F}{\partial v^2}\right|_{u=y,v=y'} \quad \text{and} \quad \frac{\partial^2 F}{\partial y \partial y'} = \left.\frac{\partial^2 F}{\partial u \partial v}\right|_{u=y,v=y'}.$$

Because you must be able to use this notation we suggest that you do all the following exercises before proceeding.

label: ex:vp2-08

Exercise 2.8

If $F(x,y') = \sqrt{x^2+y'^2}$ find $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$, $\frac{\partial F}{\partial y'}$, $\frac{dF}{dx}$ and $\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)$. Also, show that,

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

label: ex:vp2-08b

Exercise 2.9

Show that for an arbitrary differentiable function $F(x, y, y')$

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

Hence show that

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

with equality only if F does not depend explicitly upon y.

label: ex:vp2-09

Exercise 2.10

Use the first identity found in exercise 2.9 to show that the equation

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

is equivalent to the second-order differential equation

$$\frac{\partial^2 F}{\partial y'^2}y'' + \frac{\partial^2 F}{\partial y \partial y'}y' + \frac{\partial^2 F}{\partial x \partial y'} - \frac{\partial F}{\partial y} = 0.$$

Note: the first equation will later be seen as crucial to the general theory described in chapter 3. The fact that it is a second-order differential equation means that unique solutions can be obtained only if two initial or two boundary conditions are given. Also the fact that the coefficient of $y''(x)$ is $\partial^2 F/\partial y'^2$ is very important in the general theory of the existence of solutions of this type of equation.

label: ex:vp2-10

Exercise 2.11

(a) If $F(y, y') = y\sqrt{1 + y'^2}$ find $\dfrac{\partial F}{\partial y}, \dfrac{\partial F}{\partial y'}, \dfrac{\partial^2 F}{\partial y'^2}$ and show that

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = \left(1 + y'^2\right)^{-3/2}\left(y^2 \frac{d}{dx}\left(\frac{y'}{y}\right) - 1\right).$$

(b) By solving the equation $y^2(y'/y)' = 1$ show that a non-zero solution of

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0 \quad \text{is} \quad y = \frac{1}{A}\cosh(Ax + B),$$

for some constants A and B. Hint, let y be the independent variable and define a new variable z by the equation $yz(y) = dy/dx$ to obtain an expression for dy/dx that can be integrated.

2.5 Examples of functionals

label: sec:vp2-examples

In this section we describe a variety of problems that can be formulated in terms of functionals, with solutions that are stationary paths of these functionals. This list is provided because it is likely that you will not be familiar with these descriptions and will be unaware of the wide variety of problems for which variational principles are useful, and sometimes essential. You should not spend long on this section if time is short; in this case you you should aim at obtaining a rough overview of the examples. Indeed, you may move directly to chapter 3 and return to this section at a later date, if necessary.

In each of the following sub-sections a different problem is described and the relevant functional is written down; some of these are derived later. In compiling this list one aim has been to describe a reasonably wide range of applications: if you are unfamiliar with the underlying physical ideas behind any of these examples, do not worry because they are not an assessed part of the course. Another aim is to show that there are subtly different types of variational problems, for instance the isoperimetric and the catenary problems, described on pages 90 and 91 respectively.

label: sec:vp2-brach

2.5.1 The brachistochrone

Given two points $P_a = (a, A)$ and $P_b = (b, B)$ in the same vertical plane, as in the diagram below, we require the shape of the smooth wire joining P_a to P_b such that a bead sliding on the wire under gravity, with no friction, and starting at P_a with a given speed shall reach P_b in the shortest possible time.

label: f:vp2-brach

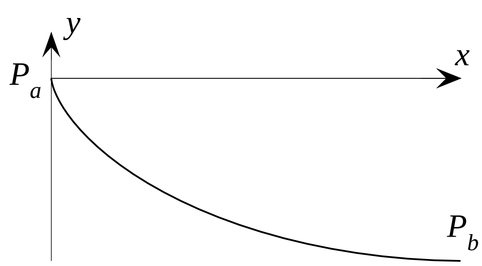

Figure 2.3 The curved line joining P_a to P_b is a segment of a cycloid. In this diagram the axes are chosen to give $a = A = 0$.

The name given to this curve is the *brachistochrone*, from the Greek, *brachystos*, shortest, and *chronos*, time.

If the y-axis is vertical it can be shown that the time taken along the curve $y(x)$ is

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

where g is the acceleration due to gravity and C a constant depending upon the initial speed of the particle. This expression is derived in section 4.2.

This problem was first considered by Galileo (1564-1642) in his 1638 work *Two New Sciences*, but lacking the necessary mathematical methods he concluded, erroneously, that the solution is the arc of a circle passing vertically through P_a; exercise 4.4 (page 156) gives part of the reason for this error.

It was John Bernoulli (1667-1748), however, who made the problem famous when in June 1696 he challenged the mathematical world to solve it. He followed his statement of the problem by a paragraph reassuring readers that the problem was very useful in mechanics, that it is not the straight line through P_a and P_b and that the curve is well known to geometers. He also stated that he would show that this is so at the end of the year provided no one else had.

In December 1696 Bernoulli extended the time limit to Easter 1697, though by this time he was in possession of Leibniz's solution, sent in a letter dated 16$^{\text{th}}$ June 1696, Leibniz having received notification of the problem on 9$^{\text{th}}$ June. Newton also solved the problem quickly: apparently[1] the letter from Bernoulli arrived at Newton's house, in London, on 29$^{\text{th}}$ January 1697 at the time when Newton was Warden of the Mint. He returned from the Mint at 4pm, set work on the problems and had solved it by the early hours of the next morning. The solution was returned anonymously, to no avail with Bernoulli stating upon receipt "The lion is recognised by his paw". Further details of this history and details of these solutions may be found in Goldstine (1980, chapter 1).

The curve giving this shortest time is a segment of a *cycloid*, which is the curve traced out by a point fixed on the circumference of a vertical circle rolling, without slipping, along a straight line. The parametric equations of the cycloid shown in figure 2.3 are

$$x = a(\theta - \sin\theta), \quad y = -a(1 - \cos\theta),$$

where a is the radius of the circle: these equations are derived in section 4.2.1, where other properties of the cycloid are discussed.

Other, historically important names, are the *isochronous* curve and the *tautochrone*. A tautochrone is a curve such that a particle travelling along it under gravity reaches a fixed point in a time independent of its starting point; a cycloid is a tautochrone besides being a brachistochrone. Isochronal means "equal times" so isochronous curves and tautochrones are the same.

There are many variations of the brachistochrone problem. Euler[2] considered the effect of resistance proportional to v^{2n}, where v is the speed and n an integer. The problem of a wire with Coulomb friction, however, was not considered until 1975[3]. Both these extensions require the use of Lagrange multipliers and are described in chapter 10. Another variation was introduced by Lagrange[4] who allowed the end point P_b, in figure 2.3 to lie on a given surface and this introduces different boundary conditions that the cycloid needs to satisfy: the simpler variant in which the motion remains in the plane and one or both end points lie on given curves is treated in chapter 8.

2.5.2 Minimal surface of revolution

Here the problem is to find a curve $y(x)$ passing through two given points $P_a = (a, A)$ and $P_b = (b, B)$, with $A \ge 0$ and $B > 0$, as shown in the diagram, such that when rotated about the x-axis the area of the curved surface formed is a minimum. label: f:vp2-minsur

[1]This anecdote is from the records of Catherine Conduitt, née Barton, Newton's niece who acted as his housekeeper in London, see *Newton's Apple* by P Aughton, 2003 (Weidenfeld and Nicolson), page 201.

[2]Chapter 3 of his 1744 opus, *The Method of Finding Plane Curves that Show Some Property of Maximum or Mininum...*

[3]Ashby A, Brittin W E, Love W F and Wyss W, 1975 *Brachitochrone with Coulomb Friction*, Amer J Physics **43** 902-5

[4]*Essay on a new method..* published in Vol II of the *Miscellanea Taurinensai*, the memoirs of the Turin Academy.

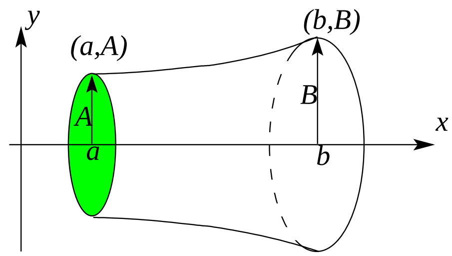

Figure 2.4 Diagram showing the cylindrical shape produced when a curve $y(x)$, joining (a, A) to (b, B), is rotated about the x-axis.

The area of this surface is shown in section 4.3 to be

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

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

2.5.3 The minimum resistance problem

label: sec:vp2-minres

An important problem in the history of the Calculus of Variations is the determination of the shape of a solid of revolution meeting with least resistance to its motion along its axis through a stationary fluid.

Newton was interested in the problem of fluid resistance and performed many experiments aimed at determining its dependence on various parameters, such as the velocity through the fluid. These experiments were described in Book II of *Principia* (1687)[5]; an account of Newton's ideas is given by Smith (2000)[6]. It is to Newton that we owe the idea of the drag coefficient, C_D, a dimensionless number allowing the force on a body moving through a fluid to be written in the form

$$F_R = \frac{1}{2} C_D \rho A_f v^2,$$

where A_f is the frontal area of the body, ρ the fluid density[7] and $v = |\mathbf{v}|$ where $\mathbf{v}$, is the relative velocity of the body and the fluid. For modern cars C_D has values between 0.30 and 0.45, with frontal areas of about 30 ft^2.

Newton distinguished two types of forces:
a) those imposed on the front of the body which oppose the motion, and
b) those at the back of the body resulting from the disturbance of the fluid and which may be in either direction.
He also considered two types of fluid:

[5] The full title is *Philopsophiae naturalis Principia Mathematica*, (Mathematical Principles of natural Philosophy.

[6] Smith G E 2000 *Fluid Resistance: Why Did Newton Change His Mind?*, in *The Foundations of Newtonian Scholarship*.

[7] Note that this suggests that the 30°C change in temperature between summer and winter changes F_R by roughly 10%.

a) *rarefied* fluids comprising non-interacting particles spread out in space, such as a gas, and
b) *continuous* fluids, comprising particles packed together so that each is in contact with its neighbours, such as a liquid.

The ideas sketched below are most relevant to rarefied fluids and ignore the second type of force. They were used by Newton in 1687 to derive a functional for which the stationary path yields, in theory, a surface of minimum resistance. This solution does not, however, agree with observation largely because the physical assumptions made are too simple, in particular the demarcation of forces into the two types listed above is unhelpful. Moreover, Weierstrass showed that this stationary path does not yield a minimum. Nevertheless, the general problem is important and Newton's approach, and the subsequent variants, are of historical and mathematical importance: we shall mention a few of these variants after describing the basic problem.

It is worth noting that the problem of fluid resistance is difficult and was not properly understood until the early part of the 20$^{\text{th}}$ century. In 1752 d'Alembert published a paper, *Essay on a New theory of the resistance of Fluids*, in which he derived the partial differential equations describing the motion of an ideal, incompressible inviscid fluid; the solution of these equations showed that resisting force was zero, regardless of the shape of the body: this was in contradiction to observations and was henceforth known as d'Alembert's paradox. It was not resolved until Prandtl (1875-1953) developed the theory of boundary layers in 1904. This shows how fluids of relatively small viscosity, such as water or air, may be treated mathematically by taking account of friction only in the region where essential, namely in the thin layer that exists in the neighbourhood of the solid body. This concept was introduced in 1904, but many decades passed before its ramifications were understood: an account of these ideas can be found in Schlichting (1955)[8] and modern account of d'Alembert's paradox can be found in Landau and Lifshitz (1959)[9].

We now return to the main problem, which is to determine a functional for the fluid resistance. In deriving this it is necessary to make some assumptions about the resistance and this, it transpires, is why the stationary path is not a minimum. The main result is given by equation 2.20, and you may ignore the derivation if you wish.

It is assumed that the resistance is proportional to the square of the velocity. To see why, consider a small plane area moving through a fluid comprising many isolated stationary particles, with density ρ: the area of the plane is δA and it is moving with velocity $\mathbf{v}$ along its normal, as seen in the left hand side of figure 2.5.

In order to derive a simple formula for the force on the area δA it is helpful to imagine the fluid as comprising many particles, each of mass m and all stationary. If there are N particles per unit volume, the density is $\rho = mN$. In the small time δt the area δA sweeps through a volume $v\delta t\delta A$, so $Nv\delta t\delta A$ particles collide with the area, as shown schematically on the left hand side of figure 2.5. label: f:vp2-a01

[8]Schlichting H *Boundary Layer Theory* (McGraw-Hill, New York).
[9]Landau L D and Lifshitz E M *Fluid mechanics* (Pergamon)

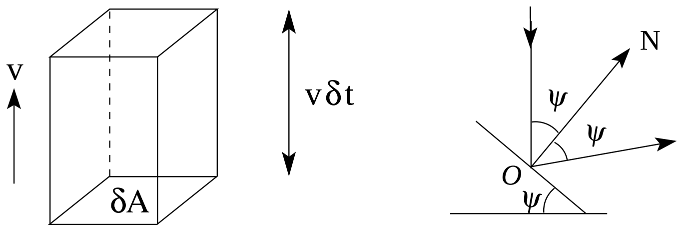

Figure 2.5 Diagram showing the motion of a small area, δA, through a rarified gas.

For an elastic collision between a very large mass (that of which δA is the small surface element) with velocity $\mathbf{v}$, and a small initially stationary mass, m, the momentum change of the light particle is $2m\mathbf{v}$ — you may check this by doing exercise 2.22, although this is not part of the course. Thus in a time δt the total momentum transfer is in the opposite direction to $\mathbf{v}$,

$$\Delta P = (2mv) \times (Nv\delta t\delta A).$$

Newton's law equates force with the rate of change of momentum, so the force on the area opposing the motion is, since $\rho = mN$,

label: eq:vp2-a01

$$\delta F = \frac{\Delta P}{\delta t} = 2\rho v^2 \delta A. \tag{2.18}$$

Equation 2.18 is a justification for the v^2-law. If the normal, ON, to the area δA is at an angle ψ to the velocity, as in the right hand side side of figure 2.5, where the arrows denote the fluid velocity relative to the body, then the formula 2.18 is modified in two ways. First, the significant area is the projection of δA onto $\mathbf{v}$, so $\delta A \to \delta A \cos\psi$. Second, the fluid particles are elastically scattered through an angle 2ψ (because the angle of incidence equals the angle of reflection), so the momentum transfer along the direction of travel is $v(1+\cos 2\psi) = 2v\cos^2\psi$: hence $2v \to 2v\cos^2\psi$, and the force in the direction $(-\mathbf{v})$ is

$$\delta F = 2\rho v^2 \cos^3\psi\, \delta A.$$

We now apply this formula to find the force on a surface of revolution. We define Oy to be the axis: consider a segment CD of the curve in the Oxy-plane, with normal PN at an angle ψ to Oy, as shown in the left-hand panel of figure 2.6.

label: f:vp2-a02

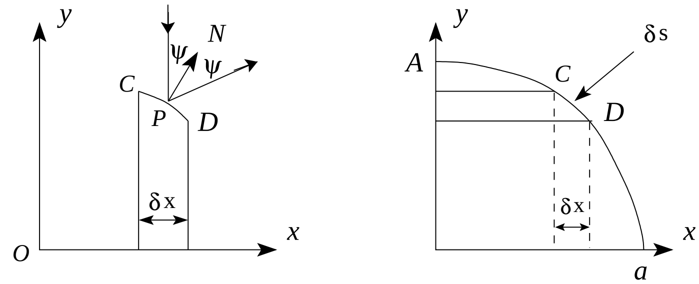

Figure 2.6 Diagram showing change in velocity of a particle colliding with the element CD, on the left, and the whole curve which is rotated about the y-axis, on the right.

The force on the ring formed by rotating the segment CD about Oy is, because of axial symmetry, in the **y**-direction. The area of the ring is $2\pi x\delta s$, where δs is the length of the element CD, so the magnitude of the force opposing the motion is

$$\delta F = 2\pi x\delta s\left(2\rho v^2\cos^3\psi\right),$$

The total force on the curve in figure 2.6 is obtained by integrating from $x=0$ to $x=a$, and is given by the functional, label: eq:vp2-a02

$$F[y] = 4\pi\rho v^2\int_{x=0}^{x=a} ds\, x\cos^3\psi, \quad y(0)=A, \quad y(a)=0. \tag{2.19}$$

But $dy/dx = \tan\psi$ and $\cos\psi = ds/dx$, so that label: eq:vp2-a03

$$F[y] = 4\pi\rho v^2\int_0^a dx\,\frac{x}{1+y'^2}, \quad y(0)=A, \quad y(a)=0. \tag{2.20}$$

For a disc of area A_f, $y'(x)=0$, and this reduces to $F=2A_f\rho v^2$. Newton's problem is to find the path making this functional a minimum.

Variations of this problem were considered by Newton: one is the curve CAB, shown in figure 2.7, rotated about Oy.

label: f:vp2-a03

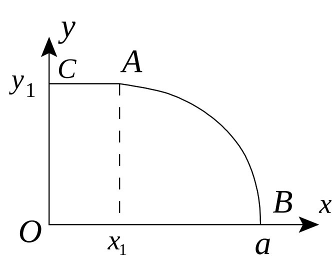

Figure 2.7 Diagram showing the modified geometry considered by Newton. Here the variable x_1 is an unkown, the line AC is parallel to the x-axis and the coordinates of C are $(0, y_1)$.

In this problem the position B is fixed, but the position of A is not; it is merely constrained to be on the line $y=y_1$, parallel to Ox and intersecting the y-axis at y_1. The resisting force is now given by the functional label: eq:vp2-a04

$$\frac{F_1[y]}{2\pi\rho v^2} = x_1^2 + 2\int_{x_1}^{a} dx\, \frac{x}{1+y'^2}, \quad y(x_1) = y_1, \quad y(a) = 0. \tag{2.21}$$

Now the path $y(x)$ *and* the number x_1 are to be chosen to make the functional a minimum.

Problems such as this, where the position of one (or both) of the end points are also to be determined are known as *variable end point problems* and are dealt with in chapter 8.

2.5.4 A problem in navigation

label:
sec:vp2-nav

Given 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.

If the y-axis is chosen to be the left bank, the starting point to be the origin and the water is assumed to be moving parallel to the banks with speed $v(x)$, a known function of the distance from the left-hand bank, then the time of passage along the path $y(x)$ is, assuming $c > \max(v(x))$,

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

where the final destination is a distance A along the right hand bank. The derivation of this result is set in exercise 2.21, one of the harder exercises at the end of this chapter.

2.5.5 The isoperimetric problem

label:
sec:vp2-isoper

Among all curves, represented by functions with continuous derivatives, that join the two points P_a and P_b in the plane and have given length L, determine that which encompasses the largest area, $S[y]$ shown in the diagram

label:
f:vp2-isop

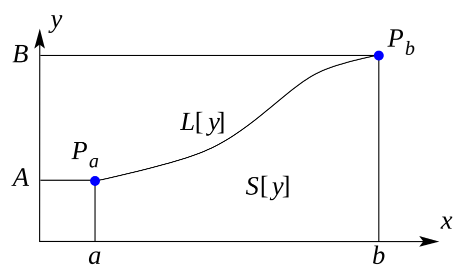

Figure 2.8 Diagram showing the area, $S[y]$, under a curve of given length joining P_a to P_b.

This is a classic problem discussed by Pappus of Alexandria in about 300 AD. Pappus showed, in Book V of his collection, that of two regular polygons having equal perimeters the one with the greater number of sides has the greater area. In the same book he demonstrates that for a given perimeter the circle has a greater area than does any regular polygon. This work seems to follow closely the earlier work of Zenodorus (circa 180 BC): extant fragments of his work include a proposition that of all solid figures, the surface areas of which are equal, the sphere has the greatest volume.

Returning to figure 2.8, a modern analytic treatment of the problem requires a differentiable function $y(x)$ satisfying $y(a) = A$, $y(b) = B$, such that the area,

$$S[y] = \int_a^b dx\, y(x)$$

is largest when the length of the curve,

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

is given. It transpires that a circular arc is the solution.

This problem differs from the first three because an additional constraint — the length of the curve — is imposed. We consider this type of problem in chapter 10.

2.5.6 The catenary

label: sec:vp2-caten

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

label: f:vp2-cat

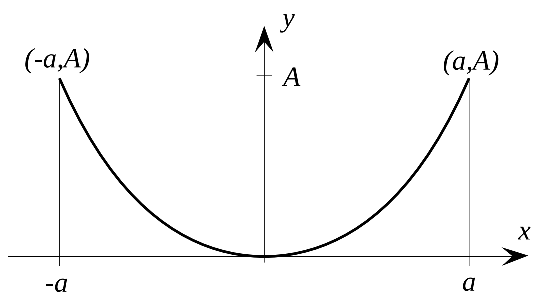

Figure 2.9 The catenary formed by a uniform chain hanging between two points at the same height.

If the lowest point of the chain is taken as the origin, the catenary equation is

label: eq:vp2-int02

$$y = c\left(\cosh\left(\frac{x}{c}\right) - 1\right) \tag{2.22}$$

for some constant c determined by the length of the chain and the value of a.

If a curve is described by a differentiable function $y(x)$ it can be shown, see exercise 2.18, that the potential energy E of the chain is proportional to the functional

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

The curve that minimises this functional, subject to the length of the chain $L = \int_{-a}^{a} dx\, \sqrt{1 + y'^2}$ remaining constant, is the shape assumed by the hanging chain. In common with the previous example, the catenary problem involves a constraint — again the length of the chain — and is dealt with in chapter 10.

2.5.7 Fermat's principle

Light and other forms of electromagnetic radiation are wave phenomena. However, in many common circumstances light may be considered to travel along lines joining the source to the observer: these lines are named *rays* and are often straight lines. This is why most shadows have distinct edges and why eclipses of the Sun are so spectacular. In a vacuum, and normally in air, these rays are straight lines and the speed of light in a vacuum is $c \simeq 2.9 \times 10^{10}$ cm/sec, independent of its colour. In other uniform media, for example water, the rays also travel in straight lines, but the speed is different: if the speed of light in a uniform medium is c_m then the refractive index is defined to be the ratio $n = c/c_m$. The refractive index usually depends on the wave length: thus for water it is 1.333 for red light (wave length 6.50×10^{-5} cm) and 1.343 for blue light (wave length 7.5×10^{-5} cm); this difference in the refractive index is one cause of rainbows. In non-uniform media, in which the refractive index depends upon the position, light rays follow curved paths. Mirages are one consequence of a position-dependent refractive index.

A simple example of the ray description of light is the reflection of light in a plane mirror. In the diagram the source is S and the light ray is reflected from the mirror at R to the observer at O. The plane of the mirror is perpendicular to the page and it is assumed that the plane SRO is in the page.

label:
f:vp2-mirr

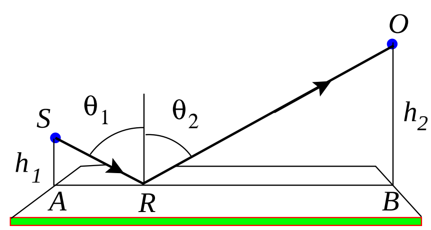

Figure 2.10 Diagram showing light travelling from a source S to an observer O, via a reflection at R. The angles of incidence and of reflection are defined to be θ_1 and θ_2, respectively.

It is known that light travels in straight lines and is reflected from the mirror at a point R as shown in the diagram. However, without further information the position of R is unknown. Observations, however, show that the angle of incidence, θ_1, and the angle of reflection, θ_2, are equal. This law of reflection was known to Euclid (circa 300 BC) and Aristotle (384-322 BC); but it was Hero of Alexandria (circa 125 BC) who showed by geometric argument that the equality of the angles of incidence and reflection is a consequence of the Aristotelean principle that nature does nothing the hard way; that is, if light is to travel from the source S to the observer O via a reflection in the mirror then it travels along the shortest path.

This result was generalised by the French mathematician Fermat (1601-1665) into what is now known as *Fermat's principle* which states[10] that the path taken by light

[10] Fermat's original statement was that light travelling between two points seeks a path such that the number of waves is equal, as a first approximation, to that in a neighbouring path. This formulation has the form of a variational principle, which is remarkable because Fermat announced this result in 1658, before the calculus of either Newton or Leibniz was developed.

rays is that which minimises the *time* of passage. For the mirror, because the speed along SR and RO is the same this means that the distance along SR plus RO is a minimum. If $AB = d$ and $AR = x$, the total distance travelled by the light ray depends only upon x and is

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

This function has a minimum when $\theta_1 = \theta_2$, that is when the angle of incidence, θ_1, equals the angle of reflection, θ_2. This result is proved in exercise 2.13.

In general, for light moving in the Oxy-plane, in a medium with refractive index $n(x,y)$, with the source at the origin and observer at (a, A) the time of passage, T, along an arbitrary path $y(x)$ joining these points is

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

This follows because the time taken to travel along an element of length δs is $n(x,y)\delta s/c$ and $\delta s = \sqrt{1+y'(x)^2}\,\delta x$. If the refractive index, $n(x,y)$, is constant then this integral reduces to the integral 2.1 and the path of a ray is a straight line, as would be expected.

Fermat's principle can be used to show that for light reflected at a mirror the angle of incidence equals the angle of reflection. For light crossing the boundary between two media it gives Snell's law,

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

where α_1 and α_2 are the angles between the ray and the normal to the boundary and c_k is the speed of light in the media, as shown in figure 2.11: in water the speed of light is approximately $c_2 = c_1/1.3$, where c_1 is the speed of light in air, so $1.3\sin\alpha_2 = \sin\alpha_1$. label: f:vp2-snell

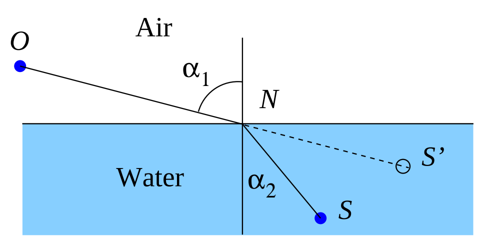

Figure 2.11 Diagram showing the refraction of light at the surface of water. The angles of incidence and refraction are defined to be α_2 and α_1 respectively; these are connected by Snell's law.

In figure 2.11 the observer at O sees an object S in a pond and the light ray from S to O travels along the two straight lines ON and NS, but the observer perceives the object to be at S', on the straight line ON. This explains why a stick put partly into water appears bent.

2.5.8 Coordinate free formulation of Newton's equations

Newton's laws of motion accurately describe a significant portion of the physical world, from the motion of large molecules to the motion of galaxies. However, Newton's original formulation is usually difficult to apply to even quite simple mechanical systems and hides the mathematical structure of the equations of motion, which is important for the advanced developments in dynamics and for finding approximate solutions. It transpires that in many important circumstances Newton's equations of motion can be expressed as a variational principle the solution of which are the equations of motion. This reformulation took some years to accomplish and was originally motivated by Snell's law and Fermat's principle, that minimises the time of passage, but also by the ancient philosophical belief in the "Economy of Nature"; for a brief overview of these ideas the introduction of the book by Yourgrau and Mandelstam (1965) should be consulted.

The first variational principle for dynamics was formulated in 1744 by Maupertuis (1698-1759), but in the same year Euler (1707-1783) described the same principle more precisely. In 1760 Lagrange (1736-1813) clarified these ideas, by first reformulating Newton's equations of motion into a form now known as Lagrange's equations of motion: these are equivalent to Newton's equations but easier to use because the form of the equations is independent of the coordinate system used — this basic property of variational principles is discussed in chapter 5 — and this allows easier use of more general coordinate systems.

The next major step was taken by Hamilton (1805-1865), in 1834, who cast Lagrange's equations as a variational principle; confusingly, we now name this Lagrange's variational principle. Hamilton also generalised this theory to lay the foundations for the development of modern physics that occurred in the early part of the 20^{th} century. These developments are important because they provide a coordinate-free formulation of dynamics which emphasises the underlying mathematical structure of the equations of motion, which is important in helping to understand how solutions behave.

Summary

These few examples provide some idea of the significance of variational principles. In summary, they are important for three distinct reasons

- A variational principle is often the easiest or the only method of formulating a problem.
- Often conventional boundary value problems may be re-formulated in terms of a variational principle which provides a powerful tool for approximating solutions.
- A variational formulation provides a coordinate free method of expressing the laws of dynamics, allowing powerful analytic techniques to be used in ordinary Newtonian dynamics. The use of variational principles also paved the way for the formulation of dynamical laws describing motion of objects moving at speeds close to that of light (special relativity), particles interacting through gravitational forces (general relativity) and the laws of the microscopic world (quantum mechanics).

2.6 Miscellaneous exercises

label: ex:vp2-01e

Exercise 2.12

Functionals do not need to have the particular form considered in this chapter. The following expressions also map functions to real numbers:

(a) $D[y] = y'(1) + y(1)^2$;

(b) $K[y] = \int_0^1 dx\, a(x) \left[y(x) + y(1)y'(x) \right]$;

(c) $L[y] = \left[xy(x)y'(x) \right]_0^1 + \int_0^1 dx \left[a(x)y'(x) + b(x)y(x) \right]$, where $a(x)$ and $b(x)$ are prescribed functions;

(d) $S[y] = \int_0^1 ds \int_0^1 dt\, \left(s^2 + st \right) y(s)y(t)$.

Find the values of these functionals for the functions $y(x) = x^2$ and $y(x) = \cos \pi x$ when $a(x) = x$ and $b(x) = 1$.

label: ex:vp2-02e

Exercise 2.13

Show that the function

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

where h_1, h_2 are defined in figure 2.10 (page 92) and x and d denote the lengths AR and AB respectively, is stationary when $\theta_1 = \theta_2$ where

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

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

label: ex:vp2-03e

Exercise 2.14

Consider the functional

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

(a) Show that the stationary function is the straight line $y(x) = Bx$ and that the value of the functional on this line is $S[y] = B\sqrt{1 + B}$.

(b) By expanding the integrand of $S[y + \epsilon h]$ to second order in ϵ, show that

$$S[y + \epsilon h] = S[y] + \frac{(4 + 3B)\epsilon^2}{8(1 + B)^{3/2}} \int_0^1 dx\, h'(x)^2, \quad B > -1,$$

and deduce that on this path the function has a minimum.

label: ex:vp2-04e

Exercise 2.15
Using the method described in the text, show that the functionals

$$S_1[y] = \int_a^b dx\, \left(1 + xy'\right) y' \quad \text{and} \quad S_2[y] = \int_a^b dx\, xy'^2,$$

where $b > a > 0$, $y(b) = B$ and $y(a) = A$ are both stationary on the same curve, namely

$$y(x) = A + (B - A)\frac{\ln(x/a)}{\ln(b/a)}.$$

Explain why the same function makes both functionals stationary.

label: ex:vp2-05e

Exercise 2.16
In this exercise the theory developed in section 2.3.1 is extended. The function $F(z)$ has a continuous second derivative and the functional S is defined by the integral

$$S[y] = \int_a^b dx\, F(y').$$

(a) Show that

$$S[y + \epsilon h] - S[y] = \epsilon \int_a^b dx\, \frac{dF}{dy'} h'(x) + \frac{1}{2}\epsilon^2 \int_a^b dx\, \frac{d^2F}{dy'^2} h'(x)^2 + O(\epsilon^3),$$

where $h(a) = h(b) = 0$.

(b) Show that if $y(x)$ is chosen to make dF/dy' constant then the functional is stationary.

(c) Deduce that this stationary path makes the functional either a maximum or a minimum, provided $F''(y') \neq 0$.

label: ex:vp2-06e

Exercise 2.17
Show that the functional

$$S[y] = \int_0^1 dx\, \left(1 + y'(x)^2\right)^{1/4}, \quad y(0) = 0, \quad y(1) = B,$$

is stationary for the straight line $y(x) = Bx$.

In addition, show that this straight line gives a minimum value of the functional only if $B < \sqrt{2}$, otherwise it gives a maximum.

Harder exercises

label: ex:vp2-22e

Exercise 2.18
If a uniform, flexible, inextensible chain of length L is suspended between two supports having the coordinates $(-b, B)$ and (a, A), with the y-axis pointing vertically upwards, show that, if the shape assumed by the chain is described by the differentiable function $y(x)$, then its length is given by $L[y] = \int_{-b}^{a} dx\, \sqrt{1 + y'^2}$ and its potential energy by

$$E[y] = g\rho \int_{-a}^{a} dx\, y\sqrt{1 + y'^2}, \quad y(-b) = B, \quad y(a) = A,$$

where ρ is the line-density of the chain and g the acceleration due to gravity.

label:
ex:vp2-23e

Exercise 2.19

This question is about the shortest distance between two points on the surface of a right-circular cylinder, so is a generalisation of the theory developed in section 2.2.

(a) If the cylinder axis coincides with the z-axis we may use the polar coordinates (ρ, ϕ, z) to label points on the cylindrical surface, where ρ is the cylinder radius. Show that the Cartesian coordinates of a point (x, y) are given by $x = \rho\cos\phi$, $y = \rho\sin\phi$ and hence that the distance between two adjacent points on the cylinder, (ρ, ϕ, z) and $(\rho, \phi + \delta\phi, z + \delta z)$ is, to first-order, given by $\delta s^2 = \rho^2\delta\phi^2 + \delta z^2$.

(b) A curve on the surface may be defined by prescribing z as a function of ϕ. Show that the length of a curve from $\phi = \phi_1$ to ϕ_2 is

$$L[z] = \int_{\phi_1}^{\phi_2} d\phi\, \sqrt{\rho^2 + z'(\phi)^2}.$$

(c) Deduce that the shortest distance on the cylinder between the two points $(\rho, 0, 0)$ and (ρ, α, ζ) is along the curve $z = \zeta\phi/\alpha$.

label:
ex:vp2-24e

Exercise 2.20

An inverted cone has its apex at the origin and axis along the z-axis. Let α be the angle between this axis and the sides of the cone, and define a point on the conical surface by the coordinates (ρ, ϕ), where ρ is the perpendicular distance to the z-axis and ϕ is the polar angle measured from the x-axis.

Show that the distance on the cone between adjacent points (ρ, ϕ) and $(\rho+\delta\rho, \phi+\delta\phi)$ is, to first-order,

$$\delta s^2 = \rho^2\delta\phi^2 + \frac{\delta\rho^2}{\sin^2\alpha}.$$

Hence show that if $\rho(\phi)$, $\phi_1 \le \phi \le \phi_2$, is a curve on the conical surface then its length is

$$L[\rho] = \int_{\phi_1}^{\phi_2} d\phi\, \sqrt{\rho^2 + \frac{\rho'^2}{\sin^2\alpha}}.$$

label:
ex:vp2-21e

Exercise 2.21

A straight river of uniform width a flows with velocity $(0, v(x))$, where the axes are chosen so the left-hand bank is the y-axis and where $v(x) > 0$. A boat can travel with constant speed $c > \max(v(x))$ relative to still water. If the starting and landing points are chosen to be the origin and (a, A), respectively, show that the path giving the shortest time of crossing is given by minimising the functional

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

label:
ex:vp2-25e

Exercise 2.22

In this exercise the basic dynamics required for the derivation of the minimum resistance functional, equation 2.20, is derived. This exercise is optional, because it requires knowledge of elementary mechanics which is not part of, or a prerequisite of, this course.

Consider a block of mass M sliding smoothly on a plane, the cross section of which is shown in figure 2.12.

label:
f:vp2-ex25e

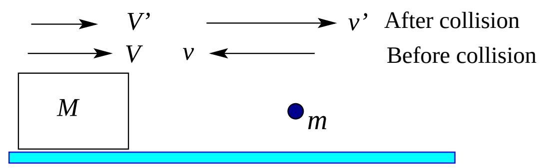

Figure 2.12 Diagram showing the velocities of the block and particle before and after the collision.

The block is moving from left to right, with speed V, towards a small particle of mass m moving with speed v, such that initially the distance between the particle and the block is decreasing. Suppose that after the inevitable collision the block is moving with speed V', in the same direction, and the particle is moving with speed v' to the right. Use conservation of energy and linear momentum to show that (V', v') are related to (V, v) by the equations

$$MV^2 + mv^2 = MV'^2 + mv'^2 \quad \text{and} \quad MV - mv = MV' + mv'.$$

Hence show that

$$V' = V - \frac{2m}{M+m}(V+v) \quad \text{and} \quad v' = \frac{2MV + (M-m)v}{M+m}.$$

Show that in the limit $m/M \to 0$, $V' = V$ and $v' = 2V + v$ and give a physical interpretation of these equations.

2.7 Solutions for chapter 2

Solution for Exercise 2.1 label: ex:vp2-02

To find the stationary function we need to compute the difference $\delta S = S[y+\epsilon g] - S[y]$ to $O(\epsilon)$, but because the second part of the question requires the second-order term we evaluate the difference to $O(\epsilon^2)$. The difference is

$$\delta S = \int_0^1 dx\, \left(\sqrt{1+y'(x)+\epsilon g'(x)} - \sqrt{1+y'(x)}\right),$$

where $g(0)=g(1)=0$. But

$$\begin{aligned}\sqrt{1+y'(x)+\epsilon g'(x)} &= \sqrt{1+y'(x)}\left(1+\frac{\epsilon g'(x)}{1+y'(x)}\right)^{1/2}, \\ &= \sqrt{1+y'(x)}\left(1+\frac{\epsilon g'(x)}{2(1+y'(x))} - \frac{\epsilon^2}{8}\left(\frac{g'(x)}{1+y'(x)}\right)^2 + \cdots\right),\end{aligned}$$

where we have used the binomial expansion $(1+z)^{1/2} = 1+\frac{1}{2}z - \frac{1}{8}z^2+\cdots$, which is equivalent to using the Taylor series for $(1+z)^{1/2}$. Hence

$$\delta S = \frac{\epsilon}{2}\int_0^1 dx\, \frac{g'(x)}{\sqrt{1+y'(x)}} - \frac{\epsilon^2}{8}\int_0^1 dx\, \frac{g'(x)^2}{(1+y'(x))^{3/2}} + O(\epsilon^3).$$

The functional is stationary if the first-order term is zero for all $g(x)$, otherwise δS would change sign with ϵ. Using the result quoted in the text (after equation 2.5) — and proved in exercise 3.4 (page 116) — this gives $\sqrt{1+y'(x)}$ =constant, that is $y'(x)$ =constant and $y(x)=\alpha x+\beta$. The boundary conditions then give $y=Bx$ for the stationary path. With this value for $y(x)$, the integrand is real if $B>-1$ and has the value $S=\sqrt{1+B}$.

Solution for Exercise 2.2 label: ex:vp2-03

(a) The required expansion is given by first writing the square root as

$$\sqrt{1+\alpha^2+2\epsilon\alpha\beta+\epsilon^2\beta^2} = \sqrt{1+\alpha^2}\left(1+\frac{2\epsilon\alpha\beta}{1+\alpha^2}+\frac{\epsilon^2\beta^2}{1+\alpha^2}\right)^{1/2}.$$

Now use the binomial expansion $(1+z)^{1/2} = 1+\frac{1}{2}z-\frac{1}{8}z^2+\cdots$ to give

$$\begin{aligned}\sqrt{1+\frac{2\epsilon\alpha\beta}{1+\alpha^2}+\frac{\epsilon^2\beta^2}{1+\alpha^2}} &= 1+\frac{1}{2}\left(\frac{2\epsilon\alpha\beta}{1+\alpha^2}+\frac{\epsilon^2\beta^2}{1+\alpha^2}\right) - \frac{1}{8}\left(\frac{2\epsilon\alpha\beta}{1+\alpha^2}+\frac{\epsilon^2\beta^2}{1+\alpha^2}\right)^2 + \cdots, \\ &= 1+\frac{\epsilon\alpha\beta}{1+\alpha^2}+\frac{\epsilon^2\beta^2}{2(1+\alpha^2)^2}+O(\epsilon^3).\end{aligned}$$

Hence

$$\sqrt{1+(\alpha+\epsilon\beta)^2} = \sqrt{1+\alpha^2}+\frac{\epsilon\alpha\beta}{\sqrt{1+\alpha^2}}+\frac{\epsilon^2\beta^2}{2(1+\alpha^2)^{3/2}}+O(\epsilon^3).$$

(b) With $\alpha = y'(x)$ and $\beta = g'(x)$ we see, using the argument described in the text, that the term $O(\epsilon)$ in the expansion of $S[y+\epsilon g]-S[y]$ is zero if $y'(x)$ =constant, hence the straight line defined by equation 2.6 makes the functional stationary. With this choice of $y(x)$, $\alpha = m$ and the second term in the above expansion gives the result quoted. The second-order term is positive for $\epsilon \neq 0$ and all $g(x)$, so the functional has a minimum along this line.

label: ex:vp2-04

Solution for Exercise 2.3

The expanson to second order in ϵ is derived in the solution to exercise 2.1(b). On the stationary path, $y = Bx$, the first order terms is, by definition, zero, so we have

$$\delta S = -\frac{\epsilon^2}{8(1+B)^{3/2}} \int_0^1 dx\, g'(x)^2 < 0, \quad B > -1.$$

Because this terms is always negative, for sufficiently small $|\epsilon|$ we have $S[y_s+\epsilon g] < S[y_s]$, where $y_s(x) = Bx$ is the stationary path, which is therefore a local maximum.

label: ex:vp2-04a

Solution for Exercise 2.4

If $a_1 = b_1 = 1$, $a_2 = z$ and $b_2 = z+u$ the three parts of the Cauchy-Schwarz inequality, page 37, are

$$\sum_{k=1}^{2} a_k^2 = 1+z^2, \quad \sum_{k=1}^{2} b_k^2 = 1+(z+u)^2, \quad \sum_{k=1}^{2} a_k b_k = 1+z^2+zu,$$

and the first result follows. There is equality only if $\mathbf{a} = \mathbf{b}$, that is $u = 0$. Divide the first inequality by $\sqrt{1+z^2}$ to derive the second result.

label: ex:vp2-05

Solution for Exercise 2.5

(a) If $F(y') = (1+y'^2)^{1/4}$ then $dF/dy' = y'/[2(1+y'^2)^{3/4}]$.

(b) If $F(y') = \sin y'$ then $dF/dy' = \cos y'$.

(c) Since $\frac{d}{dz}(e^z) = e^z$ we have $dF/dy' = F$.

label: ex:vp2-05a

Solution for Exercise 2.6

Consider the difference

$$\begin{aligned} \delta S &= S[y+\epsilon h] - S[y] = \int_a^b dx\, \Big[C(y'+\epsilon h') + D - (Cy'+D)\Big] \\ &= \epsilon C \int_a^b dx\, h'(x) = hC\Big[h(b) - h(a)\Big]. \end{aligned}$$

Since $h(a) = h(b) = 0$, $\delta S = 0$ for any $y(x)$. That is, there is no unique stationary path.

Alternatively, in this case the functional becomes

$$S[y] = \int_a^b dx\, (Cy'(x) + D) = C\,[y(b) - y(a)] + D(b-a).$$

This depends only upon C, D and the boundaries a and b: the value of the functional is therefore independent of the chosen path.

If C and D depend upon x then

$$\delta S = \epsilon \int_a^b dx\, C(x) h'(x).$$

The same theory that leads to equation 2.12 shows that $\delta S = 0$ for all $h(x)$ if and only if $C(x) =$constant, which is the case considered first. In either case there are no stationary paths.

Solution for Exercise 2.7 label: ex:vp2-06

In this example $F(x, v) = \sqrt{1 + x + v^2}$ and equation 2.16 becomes

$$v = c\sqrt{1 + x + v^2} \quad \text{where} \quad v = y'(x).$$

Squaring and rearranging this equation gives

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

Integrating this gives the solution in the form

$$y(x) - A = a \int_0^x dx \sqrt{1 + x} = \frac{2a}{3}\left((1 + x)^{3/2} - 1\right).$$

The value of a is obtained from the boundary condition $y(1) = B$, that is

$$\frac{2}{3}a = \frac{B - A}{2^{3/2} - 1} \quad \text{and hence} \quad y(x) = A + \frac{(B - A)}{(2^{3/2} - 1)}\left((1 + x)^{3/2} - 1\right).$$

Solution for Exercise 2.8 label: ex:vp2-08

If $F(x, y') = \sqrt{x^2 + y'^2}$, F is independent of y, we have

$$\frac{\partial F}{\partial y} = 0, \quad \frac{\partial F}{\partial x} = \frac{x}{\sqrt{x^2 + y'^2}} \quad \text{and} \quad \frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{x^2 + y'^2}}$$

giving

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

Since, F does not depend explicitly upon y, we have

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

and

$$\frac{\partial^2 F}{\partial x \partial y'} = -\frac{xy'}{(x^2 + y'^2)^{3/2}}, \quad \frac{\partial^2 F}{\partial y'^2} = \frac{1}{(x^2 + y'^2)^{1/2}} - \frac{y'^2}{(x^2 + y'^2)^{3/2}} = \frac{x^2}{(x^2 + y'^2)^{3/2}}$$

which gives

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

Also

$$\frac{\partial}{\partial y'}\left(\frac{\partial F}{\partial x}\right) = \frac{y''}{\sqrt{x^2+y'^2}} - \frac{(x+y'y'')y'}{(x^2+y'^2)^{3/2}} = \frac{x(xy''-y')}{(x^2+y'^2)^{3/2}},$$

so, in this case,

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

label: ex:vp2-08b

Solution for Exercise 2.9

The chain rule applied to a function $G(x, y(x), y'(x))$ has the form

$$\frac{dG}{dx} = \frac{\partial G}{\partial y'}\frac{dy'}{dx} + \frac{\partial G}{\partial y}\frac{dy}{dx} + \frac{\partial G}{\partial x}.$$

In this example, where $G = \partial F/\partial y'$, this expression becomes

$$\begin{aligned}\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) &= \frac{\partial}{\partial y'}\left(\frac{\partial F}{\partial y'}\right)\frac{dy'}{dx} + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y'}\right)\frac{dy}{dx} + \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y'}\right)\\ &= \frac{\partial^2 F}{\partial y'^2}y'' + \frac{\partial^2 F}{\partial y'\partial y}y' + \frac{\partial^2 F}{\partial x\partial y}\end{aligned}$$

which gives the required expression and is the left hand side of the inequality.

The right hand side of the inequality is

$$\begin{aligned}\frac{\partial}{\partial y'}\left(\frac{dF}{dx}\right) &= \frac{\partial}{\partial y'}\left(\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial y'}y''\right)\\ &= \frac{\partial^2 F}{\partial x\partial y'} + \frac{\partial F}{\partial y} + \frac{\partial^2 F}{\partial y\partial y'}y' + \frac{\partial^2 F}{\partial y'^2}y''\end{aligned}$$

which differs from the left hand side by the term $\partial F/\partial y$. Thus, only if F is independent of y are the derivatives equal.

label: ex:vp2-09

Solution for Exercise 2.10

Subtract the term $\partial F/\partial y$ to obtain the required result.

label: ex:vp2-10

Solution for Exercise 2.11

(a) Direct differentiation gives

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

Differentiating the second expression gives

$$\frac{\partial^2 F}{\partial y'^2} = \frac{y}{\sqrt{1+y'^2}} - \frac{yy'^2}{(1+y'^2)^{3/2}} = \frac{y}{(1+y'^2)^{3/2}}.$$

Using the expression derived in exercise 2.10, namely

$$z=\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)-\frac{\partial F}{\partial y}=y''\frac{\partial^2 F}{\partial y'^2}+y'\frac{\partial^2 F}{\partial y\partial y'}-\frac{\partial F}{\partial y}=0,\quad \text{since}\quad \frac{\partial^2 F}{\partial x\partial y'}=0,$$

we obtain

$$\begin{aligned} z &= \frac{yy''}{(1+y'^2)^{3/2}}+\frac{y'^2}{(1+y'^2)^{1/2}}-\left(1+y'^2\right)^{1/2}, \\ &= \frac{1}{(1+y'^2)^{3/2}}\left(yy''+\left(1+y'^2\right)y'^2-\left(1+y'^2\right)^2\right)=\frac{1}{(1+y'^2)^{3/2}}\left(yy''-y'^2-1\right). \end{aligned}$$

But

$$\frac{d}{dx}\left(\frac{y'}{y}\right)=\frac{y''}{y}-\frac{y'^2}{y^2}\quad\text{giving}\quad yy''-y'^2=y^2\frac{d}{dx}\left(\frac{y'}{y}\right),\quad \text{if}\quad y\neq 0,$$

and hence

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)-\frac{\partial F}{\partial y}=\frac{1}{(1+y'^2)^{3/2}}\left(y^2\frac{d}{dx}\left(\frac{y'}{y}\right)-1\right).$$

(b) If the left hand side is zero we have

$$y^2\frac{d}{dx}\left(\frac{y'}{y}\right)=1\quad\text{or}\quad y^2y'\frac{d}{dy}\left(\frac{y'}{y}\right)=1.$$

Now define $z=y'/y$ and consider z to be a function of y, so in the following $z'=dz/dy$ — note this is possible because x may be considered a function of y so y'/y can be expressed in terms of y. Now put the second equation in the form $y^3z\,z'(y)=1$, which can be integrated directly to give $z^2=C^2-y^{-2}$, for some constant C. Hence, since $z=y'/y$,

$$\frac{dy}{dx}=\sqrt{(Cy)^2-1}\quad\text{giving}\quad \int\frac{dy}{\sqrt{(Cy)^2-1}}=x+D.$$

Finally, set $Cy=\cosh\phi$ to give $\phi=C(x+D)$, that is $y=(1/C)\cosh(Cx+CD)$, which is the required solution, if $C=A$ and $CD=B$.

Solution for Exercise 2.12 label: ex:vp2-01e

(a) The expressions for $y(x)$, $y'(x)$ and $D[y]$ are

$y(x)$	$y'(x)$	$D[y]$
x^2	$2x$	3
$\cos\pi x$	$-\pi\sin\pi x$	1.

(b) If $a(x)=x$, then

$$\begin{aligned} &\text{if}\quad y(x)=x^2, && K[y]=\int_0^1 dx\,x(x^2+2x)=\frac{11}{12},\quad\text{and}\\ &\text{if}\quad y(x)=\cos\pi x, && K[y]=\int_0^1 dx\,x(\cos\pi x+\pi\sin\pi x)=-1-\frac{2}{\pi^2}. \end{aligned}$$

(c) If $a(x) = x$ and $b(x) = 1$ then

$$\text{if} \quad y(x) = x^2, \qquad L[y] = \left[2x^4\right]_0^1 + \int_0^1 dx\ \left(3x^2\right) = 3 \quad \text{and}$$

$$\text{if} \quad y(x) = \cos\pi x, \quad L[y] = \left[-\frac{\pi}{2}x\sin 2\pi x\right]_0^1 + \int_0^1 dx\ (-\pi x\sin\pi x + \cos\pi x) = -1.$$

(d) In the first case, $y(x) = x^2$,

$$\begin{aligned} S[x^2] &= \int_0^1 ds \int_0^1 dt\ \left(s^2 + st\right) s^2 t^2 = \int_0^1 ds\ \left[\frac{1}{3}s^4 t^3 + \frac{1}{4}s^3 t^4\right]_{t=0}^1 \\ &= \int_0^1 ds\ \left(\frac{1}{3}s^4 + \frac{1}{4}s^3\right) = \frac{31}{240}. \end{aligned}$$

In the second case, $y(x) = \cos\pi x$,

$$\begin{aligned} S[\cos\pi x] &= \int_0^1 ds\ \cos\pi s \int_0^1 dt\ \left(s^2 + st\right)\cos\pi t \\ &= \int_0^1 ds\ \cos\pi s\left[\frac{s^2}{\pi}\sin\pi t + s\left(\frac{t}{\pi}\sin\pi t - \frac{1}{\pi^2}\cos\pi t\right)\right]_0^1 \\ &= \frac{2}{\pi^2}\int_0^1 ds\, s\cos\pi s = -\frac{4}{\pi^4}. \end{aligned}$$

label: ex:vp2-02e

Solution for Exercise 2.13

The derivative of $f(x)$ is $f'(x) = x/\sqrt{x^2 + h_1^2} - (d-x)/\sqrt{(d-x)^2 + h_2^2}$. Since

$$\sin\theta_1 = \frac{AR}{SR} = \frac{x}{\sqrt{x^2 + h_1^2}}, \quad \text{and} \quad \sin\theta_2 = \frac{RB}{RO} = \frac{d-x}{\sqrt{(d-x)^2 + h_2^2}},$$

where the distances are defined in figure 2.10 (page 92), we see that the distance travelled by the light is stationary when $\sin\theta_1 = \sin\theta_2$, that is $\theta_1 = \theta_2$. Further since

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

the stationary point is a minimum.

label: ex:vp2-03e

Solution for Exercise 2.14

(a) We need the difference $\delta S = S[y + \epsilon g] - S[y]$ where $g(0) = g(1) = 0$, otherwise $g(x)$ is an arbitrary continuous function. Now, using the Binomial expansion

$$\sqrt{1 + \alpha + \epsilon\beta} = \sqrt{1+\alpha}\left(1 + \frac{\epsilon\beta}{2(1+\alpha)} - \frac{\epsilon^2\beta^2}{8(1+\alpha)^2} + O(\epsilon^3)\right),$$

and so

$$\begin{aligned}(\alpha+\epsilon\beta)\sqrt{1+\alpha+\epsilon\beta} &= \alpha\sqrt{1+\alpha}\left(1+\frac{\epsilon\beta}{2(1+\alpha)}-\frac{\epsilon^2\beta^2}{8(1+\alpha)^2}+\cdots\right)\\ &\quad +\epsilon\beta\sqrt{1+\alpha}\left(1+\frac{\epsilon\beta}{2(1+\alpha)}+\cdots\right),\\ &= \alpha\sqrt{1+\alpha}+\frac{\epsilon\beta(2+3\alpha)}{2\sqrt{1+\alpha}}+\frac{\epsilon^2\beta^2(4+3\alpha)}{8(1+\alpha)^{3/2}}+\cdots.\end{aligned}$$

Now substitute $\alpha = y'$ and $\beta = g'$ to obtain

$$\delta S = \epsilon\int_0^1 dx\,\frac{3+2y'}{2\sqrt{1+y'}}g'(x)+\frac{\epsilon^2}{8}\int_0^1 dx\,\frac{4+3y'}{(1+y')^{3/2}}g'(x)^2+O(\epsilon^3).$$

If $y(x)$ is a stationary path of S then the term $O(\epsilon)$ is zero. Since $g(0) = g(1) = 0$ it follows, as in the text, that $y'(x)$ =constant is a possible solution. Since $y(0) = 0$ and $y(1) = B$ this gives $y(x) = Bx$ and $S[y] = B\sqrt{1+B}$.

Alternatively, using equation 2.12 (page 80), with $F(y') = y'\sqrt{1+y'}$, we see that the stationary path is given by $F'(y')$ =constant and hence y' =constant, that is $y = mx + c$: since $y(0) = 0$ and $y(1) = B$ this gives $y(x) = Bx$.

(b) On substituting Bx for $y(x)$ we see that δS takes the value,

$$\delta S = \frac{\epsilon^2(4+3B)}{8(1+B)^{3/2}}\int_0^1 dx\,g'(x)^2+O(\epsilon^3).$$

Then, provided $B > -1$, δS is positive and the functional is a minumum on the stationary path.

Solution for Exercise 2.15 label: ex:vp2-04e

Observe that

$$S_1[y] = S_2[y]+\int_a^b dx\,y'(x) = S_2[y]+B-A.$$

That is the values of the two functionals differ by a constant, independent of the path. Hence the stationary paths of the two functionals are the same.

Consider the difference $\delta S = S_2[y+\epsilon g]-S_2[y]$ where $g(a) = g(b) = 0$:

$$\delta S = 2\epsilon\int_a^b dx\,xy'(x)g'(x)+O(\epsilon^2)$$

so that $\delta S = O(\epsilon^2)$ if $xy'(x) = c$, where c is a constant. Integrating this equation gives $y(x) = d + c\ln(x/a)$, where d is another constant. The boundary condition now give

$$A = d \quad\text{and}\quad B = d + c\ln(b/a) \quad\text{and hence}\quad y(x) = A + (B-A)\frac{\ln(x/a)}{\ln(b/a)}.$$

Solution for Exercise 2.16 label: ex:vp2-05e

(a) Consider the difference $\delta S = S[y+\epsilon g] - S[y]$ where $g(a) = g(b) = 0$, so we need the expansion

$$F(y'+\epsilon g') = F(y') + \epsilon g' \frac{dF}{dy'} + \frac{1}{2}\epsilon^2 g'^2 \frac{d^2F}{dy'^2} + \cdots .$$

Hence

$$\delta S = \epsilon \int_a^b dx\, \frac{dF}{dy'} g'(x) + \frac{1}{2}\epsilon^2 \int_a^b dx\, \frac{d^2F}{dy'^2} g'(x)^2 + O(\epsilon^3).$$

(b) If dF/dy' =constant then $\delta S = O(\epsilon^2)$ so $S[y]$ is stationary. If dF/dy' =constant then, provided $F(z)$ is not a constant or a linear function of z, $y'(x)$ is also a constant.

(c) On the stationary path $y'(x)$ is a constant and hence d^2F/dy'^2 is constant and

$$\delta S = \frac{1}{2}\epsilon^2 \frac{d^2F}{dy'^2} \int_a^b dx\, g'(x)^2 + O(\epsilon^3).$$

The integral is positive, so δS is positive or negative according as d^2F/dy'^2 is positive or negative. That is $S[y]$ is either a minimum ($d^2F/dy'^2 > 0$) or a maximum ($d^2F/dy'^2 < 0$). If $d^2F/dy'^2 = 0$ the nature of the stationary path can be determined only by expanding to higher order in ϵ.

label: ex:vp2-06e

Solution for Exercise 2.17

In this example $F(z) = (1+z^2)^{1/4}$, where we have used the notation of the previous exercise. Thus

$$F'(z) = \frac{z}{2(1+z^2)^{3/4}}, \quad F''(z) = \frac{2-z^2}{4(1+z^2)^{7/4}},$$

and hence the stationary path is $y = Bx$ and

$$S[y+\epsilon g] - S[y] = \frac{(2-B^2)\epsilon^2}{8(1+B^2)^{7/4}} \int_0^1 dx\, g'(x)^2 + O(\epsilon^3).$$

Thus if $B < \sqrt{2}$ the difference is positive for all $g(x)$ and ϵ, if sufficiently small, so the functional is a minimum along the line $f(x) = Bx$. For $B > \sqrt{2}$ the difference is negative and the functional is a maximum. If $B = \sqrt{2}$ the nature of the stationary path can be determined only by expanding to higher order in ϵ.

label: ex:vp2-22e

Solution for Exercise 2.18

The potential energy, δV, of an element of the rope of length δs centred on a point x is given by mass×height× g, that is $\delta V = (\rho\delta s)y(x)g$: since $\delta s = \sqrt{1+y'^2}\delta x$ this gives the total potential energy as $E[y] = \rho g \int_{-b}^a dx\, y\sqrt{1+y'^2}$ and $L[y] = \int_{-b}^a dx\, \sqrt{1+y'^2}$ is the length of the chain.

label: ex:vp2-23e

Solution for Exercise 2.19

(a) Since, to first-order, $\delta x = -\rho\delta\phi\sin\phi$ and $\delta y = \rho\delta\phi\cos\phi$, the distance is

$$\delta s^2 = \delta x^2 + \delta y^2 + \delta z^2 = \rho^2\delta\phi^2 + \delta z^2 = \delta\phi^2 \left(\rho^2 + \left(\frac{\delta z}{\delta\phi}\right)^2\right).$$

(b) The length along a curve is just the sum of the small elements which in the limit $\delta\phi \to 0$ becomes the integral $L[z] = \int_{\phi_1}^{\phi_2} d\phi\, \sqrt{\rho^2 + z'(\phi)^2}$.

(c) The functional $L[z]$ is the same type as that considered in section 2.3.1 hence its minimum value is given when $z(\phi)$ is a linear function of ϕ. The boundary conditions give the result quoted.

Solution for Exercise 2.20 label: ex:vp2-24e

The Cartesian coordinates of a point (ρ, ϕ) on the cone are

$$(x, y, z) = \left(\rho\cos\phi, \rho\sin\phi, \frac{\rho}{\tan\alpha}\right)$$

and for the adjacent point at $(\rho + \delta\rho, \phi + \delta\phi)$, or $(x + \delta x, y + \delta y, z + \delta z)$ in Cartesian coordinates, we have, to first order

$$\delta x = \delta\rho\cos\phi - \rho\delta\phi\sin\phi, \quad \delta y = \delta\rho\sin\phi + \rho\delta\phi\cos\phi, \quad \delta z = \frac{\delta\rho}{\tan\alpha}.$$

The distance between the two adjacent points is therefore

$$\delta s^2 = \left(1 + \frac{1}{\tan^2\alpha}\right)\delta\rho^2 + \rho^2\delta\phi^2 = \frac{\delta\rho^2}{\sin^2\alpha} + \rho^2\delta\phi^2 = \left(\rho^2 + \frac{1}{\sin^2\alpha}\left(\frac{\delta\rho}{\delta\phi}\right)^2\right)\delta\phi^2.$$

Hence the distance between the points ϕ_1 and ϕ_2 along the curve $\rho(\phi)$ is $L[\rho] = \int_{\phi_1}^{\phi_2} d\phi\, \sqrt{\rho^2 + \rho'^2\sin^{-2}\alpha}$.

Solution for Exercise 2.21 label: ex:vp2-21e

Let the velocity of the boat relative to the water be (u_x, u_y), where $c^2 = u_x^2 + u_y^2$, and we assume that u_x is positive. The velocity of the boat relative to land is therefore $(u_x, v(x) + u_y)$. If the path taken is $y(x)$ it follows that

$$\frac{dy}{dx} = \frac{u_y + v}{u_x} \quad \text{and hence} \quad u_y = u_x\frac{dy}{dx} - v.$$

Also, the time of passage is

$$T[y] = \int_0^a \frac{dx}{u_x}.$$

Now we need an expression for u_x. Since $c^2 = u_x^2 + u_y^2$, we have, on using the above expression for u_y, $(y'(x)u_x - v)^2 = c^2 - u_x^2$. This rearranges to the quadratic

$$\left(1 + y'^2\right)u_x^2 - 2vy'\,u_x - \left(c^2 - v^2\right) = 0,$$

having the solutions

$$u_x = \frac{vy' \pm \sqrt{(vy')^2 + (c^2 - v^2)(1 + y'^2)}}{1 + y'^2}.$$

Because $c > v$ this quadratic has one positive and one negative root. We need the positive root:

$$u_x = \frac{vy' + \sqrt{(vy')^2 + (c^2 - v^2)(1 + y'^2)}}{1 + y'^2} = \frac{c^2 - v^2}{\sqrt{(vy')^2 + (c^2 - v^2)(1 + y'^2)} - vy'}.$$

Hence

$$T[y] = \int_0^a dx\, \frac{\sqrt{(vy')^2 + (c^2 - v^2)(1 + y'^2)} - vy'}{c^2 - v^2} = \int_0^a dx\, \frac{\sqrt{(1 + y'^2)c^2 - v^2} - vy'}{c^2 - v^2}.$$

label: ex:vp2-25e

Solution for Exercise 2.22

The kinetic energy of a particle of mass m and velocity $\mathbf{v}$ is $\frac{1}{2}m|\mathbf{v}|^2$ and its linear momentum is $m\mathbf{v}$. For an elastic collision energy and momentum are conserved, so

$$\begin{aligned} MV^2 + mv^2 &= MV'^2 + mv'^2 && \text{Energy conservation} \\ MV - mv &= MV' + mv' && \text{Linear momentum in the direction of the block motion} \end{aligned}$$

From the second equation $v' = M(V - V')/m - v$, so conservation of energy gives

$$\begin{aligned} MV'^2 &= MV^2 + mv^2 - m\left(v - M(V - V')/m\right)^2 \\ &= MV^2 + 2Mv(V - V') - \frac{M^2}{m}(V - V')^2. \end{aligned}$$

But $V'^2 = (V - V')^2 - 2V(V - V') + V^2$ and hence

$$M\left(1 + \frac{M}{m}\right)(V - V')^2 - 2M(V + v)(V - V') = 0,$$

with solutions $V' = V$ and

$$V' = V - \frac{2m}{M + m}(V + v) \ \rightarrow\ V \text{ as } \frac{m}{M} \rightarrow 0.$$

The solution $V' = V$ gives, from the momentum equation, $v' = -v$, which is for the motion of the particle through the block and we discard this solution. The equation for v' is

$$v' = \frac{2M}{M + m}(V + v) - v = \frac{2MV - (M - m)v}{M + m} \rightarrow 2V + -v \text{ as } \frac{m}{M} \rightarrow 0.$$

When m/M is zero the solutions correspond to the elastic collision of a massless particle from a massive body when the relative velocity before and after the collision is the same.