M820

Calculus of variations and advanced calculus

Module notes and solutions

Corrected version. Version 2018.10, 19 September 2021

Derek Richards, with amendments by Ben Mestel,
assisted by colleagues and students

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.

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

First published 2008.

Second edition December 2009.

Third edition October 2018.

Edited, designed and typeset by The Open University, using LaTeX.

SUP 05390 8

3.1

Contents

Chapter 1

Introduction & Guide (E)

1.1 Introduction (E)

This module is about two related mathematical concepts which are of use in many areas of applied mathematics, are of immense importance in formulating the laws of theoretical physics and also produce important, interesting and some unsolved mathematical problems. These are the *functional* and *variational principles*, the theory of which is named *The Calculus of Variations.*

A functional is a generalisation of a function of one or more real variables. A real function of a single real variable maps an interval of the real line to real numbers: for instance, the function $(1+x^2)^{-1}$ maps the whole real line to the interval $(0,1]$; the function $\ln x$ maps the positive real axis to the whole real line. Similarly a real function of n real variables maps a domain of R^n into the real numbers.

A *functional* maps a given class of functions to real numbers. A simple example of a functional is

$$\mathrm{S}[y] \ = \int_0^1 dx\,\sqrt{1+y'(x)^2}, \quad y(0) \ = 0, \quad y(1) \ = 1, \tag{1.1}$$

which associates a real number with any real function $y(x)$ which satisfies the boundary conditions and for which the integral exists. We use the square bracket notation[1] $S[y]$ to emphasise the fact that the functional depends upon the choice of function used to evaluate the integral. In chapter 2 we shall see that a wide variety of problems can be described in terms of functionals. Notice that the boundary conditions, $y(0)=0$ and $y(1)=1$ in this example, are often part of the definition of the functional.

Real functions of n real variables can have various properties; for instance they can be continuous, they may be differentiable or they may have stationary points and local and global maxima and minima: functionals share many of these properties. In particular the notion of a stationary point of a function has an important analogy in the theory of functionals and this gives rise to the idea of a *variational principle*, which arises

[1] In this module we use conventions common in applied mathematics and theoretical physics. A function of a real variable x will usually be represented by symbols such as $f(x)$ or just f, often with no distinction made between the function and its value. Similarly, we use the older convention, $S[y]$, for a functional, to emphasise that y is itself a function.

when the solution to a problem is given by the function making a particular functional stationary. Variational principles are common and important in the natural sciences.

The simplest example of a variational principle is that of finding the shortest distance between two points. Suppose the two points lie in a plane, with one point at the origin, O, and the other at point A with coordinates $(1, 1)$, then if $y(x)$ represents a smooth curve passing through O and A the distance between O and A, along this curve is given by the functional defined in equation (1.1). The shortest path is that which minimises the value of $S[y]$. If the surface is curved, for instance a sphere or ellipsoid, the equivalent functional is more complicated, but the shortest path is that which minimises it.

Variational principles are important for three principal reasons. First, many problems are naturally formulated in terms of a functional and an associated variational principle. Several of these will be described in chapter 2 and some solutions will be obtained as the module develops.

Second, most equations of mathematical physics can be derived from variational principles. This is important partly because it suggests a unifying theme in our description of nature and partly because such formulations are independent of any particular co-ordinate system, so making the essential mathematical structure of the equations more transparent and easier to understand. This aspect of the subject is not considered in this module; a good discussion of these problems can be found in Yourgrau and Mandelstam (1968)[2].

Finally, variational principles provide powerful computational tools; we explore aspects of this theory in chapter 13.

Consider the problem of finding the shortest path between two points on a curved surface. The associated functional assigns a real number to each smooth curve joining the points. A first step to solving this problem is to find the stationary values of the functional; it is then necessary to decide which of these provides the shortest path. This is very similar to the problem of finding extreme values of a function of n variables, where we first determine the stationary points and then classify them: the important and significant difference is that the space of allowed functions is not usually finite in dimension. The infinite dimensional spaces of functions, with which we shall be dealing, has many properties similar to those possessed by finite dimensional spaces, and in the many problems the difference is not significant. However, this generalisation does introduce some practical and technical difficulties some of which are discussed in section 4.6.

In elementary calculus and analysis, the functions studied first are 'real functions, f, of one real variable', that is, functions with domain either R, or a subset of R, and codomain R. Without any other restrictions on f, this definition is too general to be useful in calculus and applied mathematics. Most functions of one real variable that are of interest in applications have smooth graphs, although sometimes they may fail to be smooth at one or more points where they have a 'kink' (fail to be differentiable), or even a break (where they are discontinuous). This smooth behaviour is related to the fact that most important functions of one variable describe physical phenomena and often arise as solutions of ordinary differential equations. Therefore it is usual to restrict attention to functions that are differentiable or, more usually, differentiable a

[2] Yourgrau W and Mandelstam S *Variational Principles in Dynamics and Quantum Theory*, Pitman.

number of times.

The most useful generalisation of differentiability to functions defined on sets other than R requires some care. It is not too hard in the case of functions of several (real) variables but we shall have to generalise differentiation and integration to *functionals*, not just to functions of several real variables.

Our presentation conceals very significant intellectual achievements made at the end of the nineteenth century and during the first half of the twentieth century. During the nineteenth century, although much work was done on particular equations, there was little systematic theory. This changed when the idea of infinite dimensional vector spaces began to emerge. Between 1900 and 1906, fundamental papers appeared by Fredholm[3], Hilbert[4], and Fréchet[5]. Fréchet's thesis gave for the first time definitions of limit and continuity that were applicable in very general sets. Previously, the concepts had been restricted to special objects such as points, curves, surfaces or functions. By introducing the concept of distance in more general sets he paved the way for rapid advances in the theory of partial differential equations. These ideas together with the theory of Lebesgue integration introduced in 1902, by Lebesgue in his doctoral thesis[6], led to the modern theory of functional analysis. This is now the usual framework of the theoretical study of partial differential equations. They are required also for an elucidation of some of the difficulties in the Calculus of Variations. However, in this introductory module, we concentrate on basic techniques of solving practical problems, because we think this is the best way to motivate and encourage further study.

1.2 Guide (E)

The principal source of information for the module is the module VLE[7] site, which you should consult regularly. In particular, there you will find the study calendar, the Tutor Marked Assignments (TMAs), their submission cut-off dates, the two specimen examinations, several screencasts and eLectures. Of particular use is the module forum (and other forums for the MSc programme) which provides an opportunity for you to discuss the module, interact with other students, and ask for help with the exercises (but not with the TMAs). The forum is moderated and is a great resource for you and for other students on the module.

The most important resources for you are these module notes with their exercises and solutions. Written by Professor Derek Richards, they have been slightly reorganised and edited, leaving most of the original notes intact. Many thanks to colleagues for their help in the production of these revised notes. Especial thanks must go to the many M820 students who have found errors in previous versions of the notes and who have made suggestions for improvements.

[3] I. Fredholm, *On a new method for the solution of Dirichlet's problem*, reprinted in Oeuvres Complètes, l'Institut Mittag-Leffler, (Malmö) 1955, pp 61-68 and 81-106.

[4] D. Hilbert published six papers between 1904 and 1906. They were republished as *Grundzüge einer allgemeinen Theorie der Integralgleichungen* by Teubner, (Leipzig and Berlin), 1924. The most crucial paper is the fourth.

[5] M. Fréchet, Doctoral thesis, *Sur quelques points du Calcul fonctionnel*, Rend. Circ. Mat. Palermo 22 (1906), pp 1–74.

[6] H. Lebesgue, Doctoral thesis, Paris 1902, reprinted in Annali Mat. Pura e Appl., 7 (1902) pp 231–359.

[7] Virtual Learning Environment, accessible via the Internet.

As Professor Richards has written, 'Very many exercises are set in the belief that mathematical ideas cannot be understood without attempting to solve problems at various levels of difficulty and that one learns most by making one's own mistakes, which is time consuming. You should not attempt all these exercise at a first reading, but these provide practice of essential mathematical techniques and in the use of a variety of ideas, so you should do as many as time permits; thinking about a problem, then looking up the solution is usually of little value until you have attempted your own solution.'

Indeed, because so many exercises have been set, they have now been classified as described in section 1.2.1.

The Calculus of Variations builds on real analysis and Appendix A introduces many of the ideas needed for our treatment of the Calculus of Variations. It is possible that you are already familiar with the mathematics described in this appendix, in which case you could start the module with chapter 2. You should ensure, however, that you have a good working knowledge of differentiation, both ordinary and partial, Taylor series of one and several variables, and differentiation under the integral sign, all of which are necessary for the development of the theory. In addition familiarity with the theory of linear differential equations with both initial and boundary value problems is assumed, although this material is revised and expanded on in Chapter 3.

1.2.1 Chapter, section and exercises labelling (E)

To assist your studies, chapters and sections have been labelled so that you can focus your studies appropriately. The key to the labelling is:

1. Label: E for Essential Core. This material forms the essential core of the module and may be included in assessment. To pass the module you should aim to achieve fluency in the essential core material.
2. Label: C for Core. This material is part of the core of the module and may be included in assessment. Although core material, it isn't quite as fundamental as the essential core material. To obtain a good pass in the module you should aim to achieve fluency in the essential core and the core material.
3. Label: B for Background Material. This material, most of which is in Appendix A, is background material which is used by the other parts of the module. You should have come across most of this subject matter before, but there may be parts that are new to you and others which you need to revise. It is suggested that at the start of your studies you take a quick look through this background material, but do not spend much time on it at the beginning, returning to it if you need it later on. The material is not assessed in itself, but you will need to be familiar with it when you do the assessments.
4. Label: O for Optional Material. This material consists of additional material/examples/topics which form an important section of the module but which are suitable for omission for students who are short of study time. Although material from these sections may be included in assessment it will not be assumed that you are familiar with the material.
5. Label: + for Extension Material. This material consists of additional material/examples/topics which provide an extension to the core and optional material.

You are not expected to study this material. Although topics from these sections may occasionally be included in assessment, it will not be assumed that you are familiar with the material.

In a similar manner the exercises are labelled as follows:

1. Label: E for Essential. These exercises are essential as they give essential practice in the core material and are highly useful practice for module assessment. Full solutions are given to these exercises.
2. Label: R for Recommended. These exercises are recommended as they give practice in the core material and are useful practice for module assessment. Full solutions are given to these exercises.
3. Label: B for Background. These exercises give practice in the background material and they should be studied as needed. Most of these exercises are in Appendix A.
4. Label: O for Optional. These exercises either provide practice in optional material or are optional exercises for core material. Solutions may only be in outline, with gaps to be filled by the reader. The study of these exercises is not required for module assessment. Although these exercises may be used for assessment, it will not be assumed that you are familiar with the material.
5. Label: + for Extension. These exercises provide practice and additional material/examples/topics which provide an extension to the core and/or optional material. Solutions may only be in outline, with gaps to be filled by the reader. The study of these exercises is not required for module assessment. Although these exercises may be used for assessment, it will not be assumed that you are familiar with the material.

1.2.2 Overview of the chapters (E)

Chapter 2. The Calculus of Variations (E) This chapter gives an introduction to the principal ideas of the Calculus of Variations and gives a description of the main applications of the theory that are discussed later on in the notes. The applications are important historically and interesting intrinsically, but you will not be required to reproduce them. However, there may well be applications as part of the module assessment.

Chapter 3. Ordinary differential equations (E) This is an important chapter as much of the subsequent material requires a facility with the techniques of solving ordinary differential equations described in this chapter. You may have met much of the material in this chapter before. You should study this chapter, but do not spend too long at the first reading. It is better to return to the chapter as needed when you are studying the rest of the notes. The most important sections of chapter 3 are: 3.1 Order notation; 3.3 General definitions; 3.4 First-order equations, in particular 3.4.2 Separable and homogeneous equations; 3.4.3 Linear first-order equations; 3.4.5 Riccati's equation; and 3.5 Second-order equations.

Chapter 4. The Euler–Lagrange equation (E) The Euler–Lagrange equation is fundamental to the calculus of variations and you must study this chapter in particular detail. After reading the introductory material, you must understand the Gâteaux differential of a functional and what it means for a functional to be stationary (4.2.3). The sections 4.3 Fundamental Lemma; 4.4 The Euler–Lagrange Equations; and 4.4.1

The first-integral are particularly important. You might find the section 4.6 on Strong and Weak Extrema initially puzzling, but it is important to understand the difference between them.

Chapter 5. Applications of the Euler–Lagrange Equation (E) This chapter applies the Euler–Lagrange equation to some of the problems introduced in Chapter 2. The material is very interesting but not easy. The brachistochrone is one of the most famous examples in the Calculus of Variations and is particularly worthy of study.

Chapter 6. Further theoretical developments (E) This chapter contains important material, in particular 6.2 Invariance of the Euler–Lagrange equation and 6.3 Functionals with many dependent variables.

Chapter 7. Symmetries and Noether's theorem (E) Many mathematical problems which would otherwise be difficult can be solved more easily if we may make use of any symmetries. This important chapter shows that symmetries of functionals lead to an invariant first-integral through Noether's theorem.

Chapter 8. The second variation (E) In this chapter the second variation of a functional is analysed and conditions are derived for a stationary solution to be a local weak maximum or a local weak minimum. You may already be familiar with the results of subsection 8.2 Stationary points of functions of several variables. Particularly important material in this chapter is as follows. In 8.3 the second variation of a functional is defined and in 8.4 the second variation is analysed to get a sufficient criterion for a local extremum in terms of the non-existence of conjugate points and the Jacobi equation. This theory is applied to the Brachistochrone problem in 8.6.

Chapter 9. The parametric representation of functionals (E) In this chapter, functionals defined in terms of curves that are expressed parametrically are discussed. In 9.1 the parametric representation of curves is discussed and in 9.2 parametrically defined functionals are introduced, and a homogeneity criterion is given for a given parametric functional to be a representation of a standard functional. An important application is to geodesic curves on surfaces (9.2.1).

Chapter 10. Variable end points (E) This chapter describes how to deal with variational problems in which one of the endpoints is not fixed, but replaced by a natural boundary condition. Important sections are: 10.1 Introduction; 10.2 Natural boundary conditions; 10.3 Variable end points; 10.4 Parametric functionals.

Chapter 11. Conditional stationary points (E) In this chapter we discuss the method of Lagrange multipliers to find extrema of functions of several variables subject to a constraint. This material is sometimes included in undergraduate mathematics degrees, but is important so it is reviewed in this module.

Chapter 12. Constrained variational problems (E) In this chapter, we introduce functionals for which the stationary curves are subject to a constraint, usually also expressed in functional form. These are studied by forming an auxiliary functional incorporating the constraint via a Lagrange multiplier. The catenary studied in 12.2.3 is a classical application of this theory. Section 12.6 on the Lagrange Problem is not assessed. The sections 12.7 and 12.8 are interesting but lengthy extensions of the brachistochrone problem.

Chapter 13. Sturm–Liouville systems (C) The material covered in Chapter 13 is an important area of mathematics in its own right, forming the foundation of the solution of many of the classical partial differential equations of mathematical physics.

This topic in advanced calculus is connected with the Calculus of Variations, although much of the material in this chapter can be studied separately from the main theme of the module. It is therefore designated as core rather than essential material. The sections of particular importance are: 13.1 Introduction; 13.2 The Origin of Sturm–Liouville Systems; 13.3 Eigenvalues and functions of simple systems (apart from the optional 13.3.1 on Bessel functions); 13.4 Sturm–Liouville Systems (except the optional 13.4.3 Oscillation theorem). Note that section 13.2 The Origin of Sturm–Liouville Systems is not assessed.

Chapter 14. The Rayleigh–Ritz method (C) The final chapter of the module combines the work in Chapters 13 Sturm–Liouville systems and 12 Constrained variational problems to introduce an important approximation technique: the Rayleigh–Ritz method. It is used when (as is frequently the case) it is not possible to obtain an exact solution of a Sturm–Liouville problem.

Appendix A. Background material on calculus (B) This chapter contains preparatory and revision material which you will most likely have encountered previously. Make sure you are familiar with all of this material. Of particular importance are the sections on partial derivatives, implicit functions, Taylor series for several variables and integration.

Appendix B. Solutions to Exercises This appendix includes the solutions to all the exercises in the module notes. Some solutions are full and others are concise or in outline only.

Chapter 2

The Calculus of Variations (E)

2.1 Introduction (E)

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, 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 4 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 module.

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 differences between infinite and finite dimensional spaces which are not considered in this module.

2.2 The shortest distance between two points in a plane (E)

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

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

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 4, using analogies with the theory of functions of many real variables.

2.2.1 The stationary distance (E)

In the theory of functions of several real variables a stationary point is 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, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, 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, equation (A.35) (page 373), gives,

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

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 $k = 1, 2, \ldots, n$.

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

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

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

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

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,

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

Substituting this expansion into the integral and rearranging gives the difference between the two lengths,

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

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

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

for all suitable functions $h(x)$.

We shall see in chapter 4 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

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

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 4.3 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 :

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

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.

Exercise 2.1 (E) 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 (E)

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 8 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 point 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.*

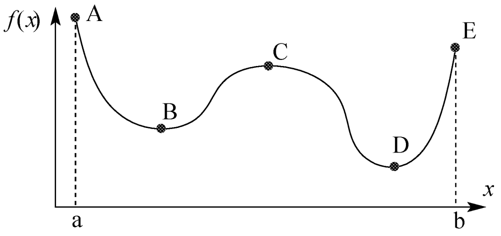


Figure 2.1: Diagram to illustrate the difference between local and global extrema.

It is clear from this example that to classify a point as a local extremum requires an examination of the function values only in the neighbourhood of the point. Whereas, determining whether a point is a global extremum 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 $\mathbf{x} = \mathbf{a}$,

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

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

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

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 8.

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,

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

for some quantity $\Delta_2[y, h]$, depending upon both y and h; special cases of this expansion are found in exercises 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 8. For the functional defined by equation (2.1), however, the proof is straightforward; the following exercise guides you through it.

Exercise 2.2 (R)

(a) Use the binomial expansion, exercise A.28 (page 372), to obtain the following expansion in ϵ,

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

(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,

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

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

Exercise 2.3 (E) 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

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

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

Now we show that the straight line between the points $(0, 0)$ and (a, A) gives a *global* minimum of the functional, not just a local minimum. This analysis relies on a special property of the integrand that follows from the Cauchy–Schwarz inequality.

Exercise 2.4 (E) Use the Cauchy–Schwarz inequality (page 378) with $\mathrm{a} = (1, z)$ and $\mathrm{b} = (1, z + u)$ 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

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

The distance between the points $(0, 0)$ and (a, A) along the path $y(x)$ is

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

On using the inequality derived in the previous exercise, with $z = y'(x)$ and $u = h'(x)$, we see that

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

But on the stationary path y' is a constant and since $h(0) = h(a) = 0$ we have $S[y + h] \geq S[y]$ for all $h(x)$.

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

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 5.20 (page 135). 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[1]. 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 (O)

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

[1] In some texts the name geodesic is used only for the shortest path.

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.

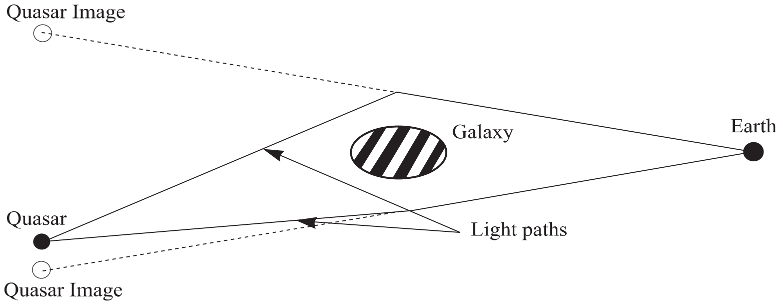


Figure 2.2: Diagram showing how an intervening galaxy can sufficiently distort a path of light from a bright object, such as a quasar, to provide two stationary paths and hence two images. Many examples of such multiple images, and more complicated but similar optical effects, have now been observed.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

2.3 Two generalisations (E)

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

The functional (2.1) (page 17) 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.

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

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

where $F(y')$ simply means that in $F(z)$ all occurrences of z are replaced by $y'(x)$. Thus for the distance between two points $F(z) = \sqrt{1+z^2}$ 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

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

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

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

where the notation dF/dy' means

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

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

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

Exercise 2.5 (E) Find the expressions for dF/dy' when

(a) $F\left(y'\right)=\left(1+y'^2\right)^{1/4}$,

(b) $F\left(y'\right)=\sin y'$,

(c) $F\left(y'\right)=\exp\left(y'\right)$. □

Substituting the difference (2.9) into the equation (2.8) gives

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

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 unknown function that we need to find. Observe that if

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

then

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

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 always a straight line, independent of the form of the integrand, although its nature can sometimes depend upon the boundary conditions, see for instance exercise 2.18 (page 40).

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.

Exercise 2.6 (O) 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 $y'(x)=\text{constant}$.

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

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

Now consider the slightly more general functional

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

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

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

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

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

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

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

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

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

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

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

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.

Exercise 2.7 (R) 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

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

2.4 Notation (E)

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

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

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 (A.20) (page 364)

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

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.

Exercise 2.8 (E) 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,

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

□

Exercise 2.9 (R) 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. □

Exercise 2.10 (R) 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 4. 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. Note also that the coefficient of $y''(x)$, $\partial^2 F/\partial y'^2$, is very important in the general theory of the existence of solutions of this type of equation.

Exercise 2.11 (O) (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 the equation

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0 \quad \text{becomes} \quad y\frac{d^2y}{dx^2} - 1 - \left(\frac{dy}{dx}\right)^2 = 0$$

and also 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

$$y\frac{d^2y}{dx^2} - 1 - \left(\frac{dy}{dx}\right)^2 = 0 \quad \text{is} \quad y = \frac{1}{A}\cosh\left(Ax+B\right),$$

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 (E)

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 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 module. Another aim is to show that there are subtly different types of variational problems, for instance the isoperimetric and the catenary problems, described in sections 2.5.5 and 2.5.6 respectively.

2.5.1 The brachistochrone (E)

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.

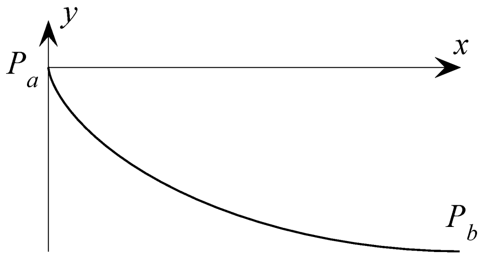


Figure 2.3: The curved line joining P_a to P_b is a segment of a cycloid. In this diagram the axes are chosen to give $a = A = 0$.

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

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

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

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 5.2.

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 5.4 (page 118) 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 16th June 1696, Leibniz having received notification of the problem on 9th June. Newton also solved the problem quickly: apparently[2] the letter from Bernoulli arrived at Newton's house, in London, on 29th January 1697 at the time when Newton was Warden of the Mint. He returned from the Mint at 4 pm, set to 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 A).

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

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

where a is the radius of the circle: these equations are derived in section 5.2.1, 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 and a brachistochrone. Isochronal means "equal times" so isochronous curves and tautochrones are the same.

There are many variations of the brachistochrone problem. Euler[3] considered the effect of resistance proportional to v^{2n}, where v is the speed and n an integer. The problem of a wire with friction, however, was not considered until 1975[4]. Both these extensions require the use of Lagrange multipliers and are described in chapter 11. Another variation was introduced by Lagrange[5] 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 10.

2.5.2 Minimal surface of revolution (E)

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 $A \geq 0$ 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.

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

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

[2] 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, Weidenfeld and Nicolson, page 201.

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

[4] Ashby A, Brittin W E Love, W F and Wyss W, *Brachistochrone with Coulomb Friction*, Amer J Physics **43** 902–5.

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

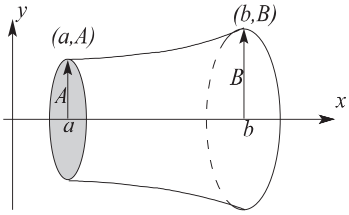


Figure 2.4: Diagram showing the cylindrical shape produced when a curve $y(x)$, joining (a, A) to (b, B), is rotated about the x-axis.

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 (C)

Newton formulated one of the first problems to involve the ideas of the Calculus of Variations. Newton's problem is to determine the shape of a solid of revolution with the 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)[6]; an account of Newton's ideas is given by Smith (2000)[7]. 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

$$F_R = \frac{1}{2} C_D \rho A_f v^2, \tag{2.18}$$

where A_f is the frontal area of the body, ρ the fluid density,[8] $v = |\mathbf{v}|$ where $\mathbf{v}$ is the relative velocity of the body and the fluid. For modern cars C_D has values between about 0.30 and 0.45, with frontal areas of about 30 ft^2 (about 2.8 m^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:

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.

[6] The full title is *Philosophiae Naturalis Principia Mathematica*, (Mathematical Principles of Natural Philosophy).

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

[8] Note that this suggests that the 30°C change in temperature between summer and winter changes F_R by roughly 10%. The density of dry air is about 1.29 kg m^{-3}.

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, equation (2.21) below, 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. Moreover, the functional has no continuously differentiable paths that can satisfy the boundary conditions, although stationary paths with one discontinuity in the derivative exist; but, Weierstrass (1815–1897) showed that this path does not yield a strong minimum. These details are discussed further in section 10.6. 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 20th century. In 1752 d'Alembert, (1717–1783), 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 the 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)[9] and a modern account of d'Alembert's paradox can be found in Landau and Lifshitz (1959)[10]. An effect of the boundary layer, and also turbulence, is that the drag coefficient, defined in equation (2.18), becomes speed dependent; thus for a smooth sphere in air it varies between 0.07 and 0.5, approximately.

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 (2.21), 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 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.

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 2.23, although this is not part of the module. Thus in a time δt the total momentum

[9] Schlichting H *Boundary Layer Theory*, McGraw-Hill.

[10] Landau L D and Lifshitz E M *Fluid mechanics*, Pergamon.

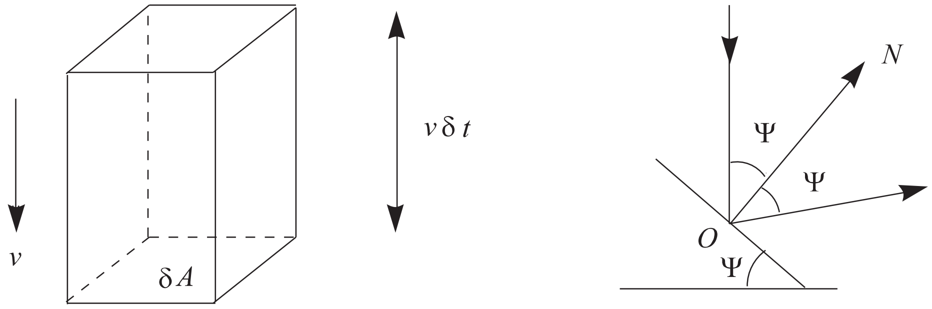


Figure 2.5: Diagram showing the motion of a small area, δA, through a rarefied gas. On the left-hand side the normal to the area is perpendicular to the relative velocity; on the right-hand side the area is at an angle. The direction of the arrows is in the direction of the gas velocity relative to the area.

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 $\rho = mN$,

$$\delta F = \frac{\Delta P}{\delta t} = 2\rho v^2 \delta A. \tag{2.19}$$

Equation (2.19) 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 of figure 2.5, where the arrows denote the fluid velocity relative to the body, then the formula (2.19) 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.

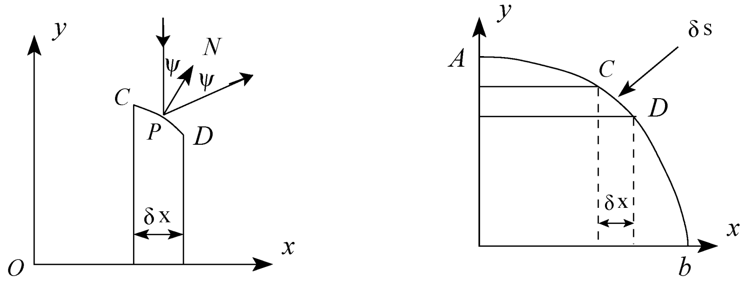


Figure 2.6: Diagram showing change in velocity of a particle colliding with the element CD, on the left, and the whole curve which is rotated about the y-axis, on the right.

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 $x = b$, and is given by the functional,

$$F[y] = 4\pi\rho v^2 \int_{x=0}^{x=b} ds\, x\cos^3\psi, \quad y(0) = A, \quad y(b) = 0. \tag{2.20}$$

But $dy/dx = \tan\psi$ and $\cos\psi = dx/ds$, so that

$$\frac{F[y]}{4\pi\rho v^2} = \int_0^b dx\, \frac{x}{1+y'^2}, \quad y(0) = A, \quad y(b) = 0. \tag{2.21}$$

For a disc of area A_f, $y'(x) = 0$, and this reduces to $F = 2A_f\rho v^2$, giving a drag coefficient $C_D = 4$, which compares with the measured value of about 1.3. Newton's problem is to find the path making this functional a minimum and this is solved in section 10.6.

Exercise 2.12 (O) Use the definition of the drag coefficient, equation (2.18), to show that, according to the theory described here,

$$C_D = \frac{8}{b^2}\int_0^b dx\, \frac{x}{1+y'^2}.$$

Show that for a sphere, where $x^2 + y^2 = b^2$ this gives $C_D = 2$. The experimental value of the drag coefficient for the motion of a sphere in air varies between 0.07 and 0.5, depending on its speed. □

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

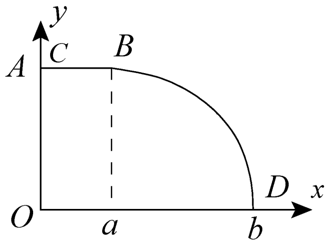


Figure 2.7: Diagram showing the modified geometry considered by Newton. Here the variable a is an unknown, the line CB is parallel to the x-axis and the coordinates of C are $(0, A)$.

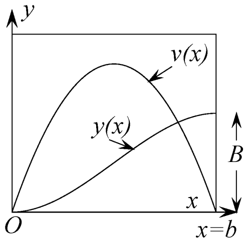
In this problem the position D is fixed, but the position of B is not; it is merely constrained to be on the line $y = A$, parallel to Ox. The resisting force is now given by the functional

$$\frac{F_1[y]}{4\pi\rho v^2} = \frac{1}{2}a^2 + \int_a^b dx\, \frac{x}{1+y'^2}, \quad y(a) = A, \quad y(b) = 0. \tag{2.22}$$

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

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 10.

2.5.4 A problem in navigation (O)



Given a river with straight, parallel banks a distance b 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, O, 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 $c > \max(v(x))$,

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

where the final destination is a distance B along the right-hand bank. The derivation of this result is set in exercise 2.22, one of the harder exercises at the end of this chapter.

A variation of this problem is obtained by not defining the terminus, so there is only one boundary condition, $y(0) = 0$, and then we need to find both the path, $y(x)$ and the terminal point. It transpires that this is an easier problem and that the path is the solution of $y'(x) = v(x)/c$, as is shown in exercise 10.7 (page 222).

2.5.5 The isoperimetric problem (E)

Among all curves, $y(x)$, represented by functions with continuous derivatives, that join the two points P_a and P_b in the plane and have given length $L[y]$, determine that which encompasses the largest area, $S[y]$, shown in diagram 2.8.

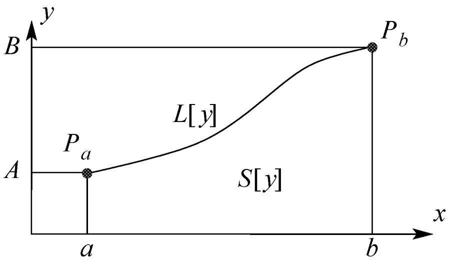


Figure 2.8: Diagram showing the area, $S[y]$, under a curve of given length joining P_a to P_b.

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,

$$S[y] = \int_a^b dx\; y$$

is largest when the length of the curve,

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

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 12.

2.5.6 The catenary (E)

A catenary is the shape assumed by an inextensible cable, or chain, of uniform density hanging between supports at both ends. In figure 2.9 we show an example of such a curve when the points of support, $(-a, A)$ and (a, A), are at the same height.

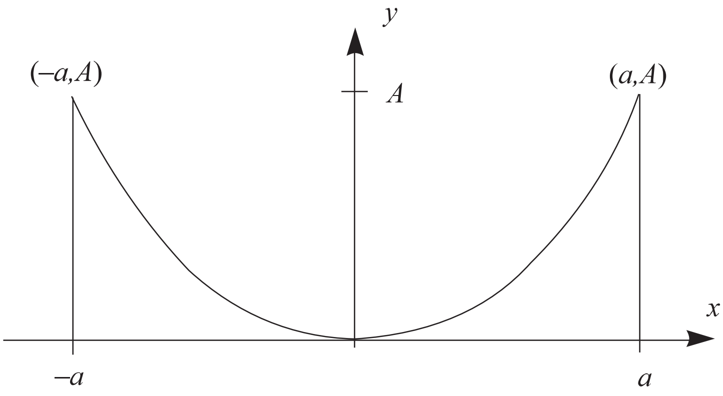


Figure 2.9: Diagram showing the catenary formed by a uniform chain hanging between two points at the same height.

If the lowest point of the chain is taken as the origin, the catenary equation is shown in section 12.2.3 to be

$$y = c\left(\cosh\left(\frac{x}{c}\right) - 1\right) \tag{2.23}$$

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 2.19, 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 $L[y] = \int_{-a}^{a} dx\; \sqrt{1 + y'^2}$ 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 using the methods described in chapter 12.

2.5.7 Fermat's principle (C)

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 $c \simeq 3.0 \times 10^{10}$ 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 6.5×10^{-5} cm) and 1.343 for blue light (wave length 4.5×10^{-5} cm); this difference in the refractive index is one cause of rainbows. In non-uniform media, in which the refractive index depends upon 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 figure 2.10 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.

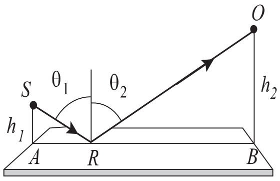


Figure 2.10: Diagram showing light travelling from a source S to an observer O, via a reflection at R. The angles of incidence and of reflection are defined to be θ_1 and θ_2, respectively.

It is known that light travels in straight lines and is reflected from the mirror at a point R as shown in the diagram. But 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 that the path taken by light rays is that which minimises the *time* of passage[11]. 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, see exercise 2.14.

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 $1.3 \sin\alpha_2 = \sin\alpha_1$.

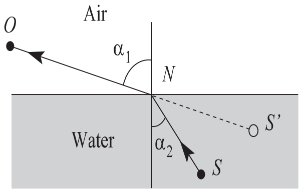


Figure 2.11: Diagram showing the refraction of light at the surface of water. The angles of incidence and refraction are defined to be α_2 and α_1 respectively; these are connected by Snell's law.

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 SN and NO, but the observer perceives the object to be at S', on the straight line OS'. This explains why a stick put partly into water appears bent.

[11] 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.

2.5.8 Coordinate free formulation of Newton's equations (C)

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 is the equations of motion. This reformulation took some years to accomplish and was originally motivated partly by Snell's law and Fermat's principle, that minimises the time of passage, and partly 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 (1968) 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 6 — 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 20th 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.

This technique is introduced in chapter 13.

- 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

Exercise 2.13 (O) 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) $K[y] = \displaystyle\int_0^1 dx\ a(x)[y(x) + y(1)y'(x)]$;

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

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

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

Exercise 2.14 (O) 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 36) 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. □

Exercise 2.15 (E) 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 functional has a minimum. □

Exercise 2.16 (E) Using the method described in the text, show that the functionals

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

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. □

Exercise 2.17 (O) 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

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

(a) Show that

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

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 $F''(y') \neq 0$. □

Exercise 2.18 (E) Show that the functional

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

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

Exercise 2.19 (O) If a uniform, flexible, inextensible chain of length L is suspended between two supports having the coordinates (a, A) and (b, B), 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 $L[y] = \int_a^b dx\, \sqrt{1 + y'^2}$ and its potential energy by

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

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

Exercise 2.20 (E) 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, $(\rho, \phi,\ z)$ 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 $z = \zeta\phi/\alpha$. □

Exercise 2.21 (E) 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

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

Exercise 2.22 (O) A straight river of uniform width b 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 (b, B), respectively, show that the path giving the shortest time of crossing is given by minimising the functional

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

Exercise 2.23 (O) In this exercise the basic dynamics required for the derivation of the minimum resistance functional, equation (2.21), is derived. This exercise is optional, because it requires knowledge of elementary mechanics which is not part of, or a prerequisite of, this module.

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

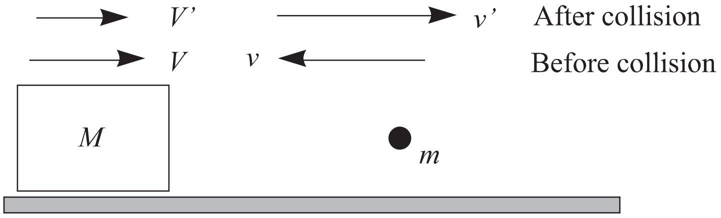


Figure 2.12: Diagram showing the velocities of the block and particle before and after the collision.

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. □

B.1 Solutions for chapter 2

Solution to Exercise 2.1

To find the stationary function we need to compute the difference $\delta S = S[y+\epsilon h] - S[y]$ to $O(\epsilon)$ but, because exercise 2.3 requires the second-order term, we evaluate the difference to $O(\epsilon^2)$. The difference is

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

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

$$\sqrt{1+y'(x)+\epsilon h'(x)} = \sqrt{1+y'(x)}\left(1+\frac{\epsilon h'(x)}{1+y'(x)}\right)^{1/2},$$

$$= \sqrt{1+y'(x)}\left(1+\frac{\epsilon h'(x)}{2\,(1+y'(x))} - \frac{\epsilon^2}{8}\left(\frac{h'(x)}{1+y'(x)}\right)^2 + \dots\right),$$

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

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

The functional is stationary if the first-order term is zero for all $h(x)$, otherwise δS would change sign with ϵ. Using the result quoted in the text (after equation (2.5)) — and proved in exercise 4.4 (page 94) — 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 to Exercise 2.2

(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 $(1+z)^{1/2} = 1+\frac{1}{2}z-\frac{1}{8}z^2+\dots$ to give

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

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

Hence

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

(b) With $\alpha = y'(x)$ and $\beta = h'(x)$ we see, using the argument described in the text, that the term $O(\epsilon)$ in the expansion of $S[y+\epsilon h] - S[y]$ 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 $h(x)$, so the functional has a minimum along this line.

Solution to Exercise 2.3

The expansion to second-order in ϵ is derived in the solution to exercise 2.1. On the stationary path, $y = Bx$, the first-order term is, by definition, zero, so we have

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

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

Solution to 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 378) 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.

Solution to 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 $\dfrac{d}{dz}(e^z) = e^z$ we have $dF/dy' = F$.

Solution to Exercise 2.6

Consider the difference

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

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. When C is not a constant there are no stationary paths.

Solution to Exercise 2.7

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

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

Solution to Exercise 2.8

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

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

Also

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

so, in this case, $\dfrac{d}{dx}\left(\dfrac{\partial F}{\partial y'}\right) = \dfrac{\partial}{\partial y'}\left(\dfrac{dF}{dx}\right)$.

Solution to 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.

Solution to Exercise 2.10

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

Solution to Exercise 2.11

(a) Direct differentiation gives $\dfrac{\partial F}{\partial y} = \sqrt{1+y'^2}$, $\dfrac{\partial F}{\partial y'} = \dfrac{yy'}{\sqrt{1+y'^2}}$. Differentiating the second expression gives

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

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

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

hence the equation $z = 0$ becomes $yy'' - 1 - y'^2 = 0$. 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 $y^3zz'(y) = 1$, which can be integrated directly to give $z^2 = C^2 - y^{-2}$, for some constant C. Hence, since $z = y'/y$, $\dfrac{dy}{dx} = \sqrt{(Cy)^2 - 1}$ giving $\displaystyle\int \frac{dy}{\sqrt{(Cy)^2-1}} = x + D$. 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 to Exercise 2.12

The first result follows directly by replacing $F[y]$, in equation (2.21), by F_R from equation (2.18). Putting $x = b\cos\theta$ and $y = b\sin\theta$ in the integral we obtain,

$$C_D = 8\int_0^{\pi/2} d\theta\ \sin^3\theta\cos\theta = 2.$$

Solution to Exercise 2.13

(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

$$\text{if } y(x) = x^2, \quad K[y] = \int_0^1 dx\ x(x^2 + 2x) = \frac{11}{12} \quad \text{and}$$

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

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

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

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

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

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

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

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

Solution to Exercise 2.14

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 36), 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

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

the stationary point is a minimum.

Solution to Exercise 2.15

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

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

and so

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

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

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

If $y(x)$ is a stationary path of S then the term $O(\epsilon)$ is zero. Since $h(0)=h(1)=0$ it follows, as in the text, that $y'(x) = \text{constant}$ 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 24), with $F(y') = y'\sqrt{1+y'}$, we see that the stationary path is given by $F'(y') = \text{constant}$ and hence $y' = \text{constant}$, 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,

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

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

Solution to Exercise 2.16

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 $\delta S = S_2[y+\epsilon h]-S_2[y]$ where $h(a)=h(b)=0$:

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

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 conditions 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 to Exercise 2.17

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

$$F\left(y' + \epsilon h'\right) = F\left(y'\right) + \epsilon h' \frac{dF}{dy'} + \frac{1}{2}\epsilon^2 h'^2 \frac{d^2F}{dy'^2} + \dots$$

Hence

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

(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

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

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.

Solution to Exercise 2.18

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

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

and hence the stationary path is $y = Bx$, $B > 0$, and

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

Thus if $B < \sqrt{2}$ the difference is positive for all $h(x)$ 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.

Solution to Exercise 2.19

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 $E[y] = \rho g \int_a^b dx\, y\sqrt{1+y'^2}$ and $L[y] = \int_a^b dx\, \sqrt{1+y'^2}$ is the length of the chain.

Solution to Exercise 2.20

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

(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 to Exercise 2.21

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

Solution to Exercise 2.22

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

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

Solution to Exercise 2.23

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

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

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

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

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

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

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.