Applied Partial Differential Equations

REVISED EDITION

$$a\left[\frac{\partial^2 u}{\partial x^2}\right]_-^+ + b\left[\frac{\partial^2 u}{\partial x \partial y}\right]_-^+ + \frac{\partial a}{\partial x}\left[\frac{\partial u}{\partial x}\right]_-^+ + \frac{\partial b}{\partial x}\left[\frac{\partial u}{\partial y}\right]_-^+ = \alpha\left[\frac{\partial u}{\partial x}\right]_-^+$$

JOHN OCKENDON | SAM HOWISON | ANDREW LACEY | ALEXANDER MOVCHAN

OXFORD

Applied Partial Differential Equations

Applied Partial Differential Equations

REVISED EDITION

JOHN OCKENDON

Oxford Centre for Industrial and Applied Mathematics
University of Oxford

SAM HOWISON

Oxford Centre for Industrial and Applied Mathematics
University of Oxford

ANDREW LACEY

Department of Mathematics
Heriot-Watt University

and

ALEXANDER MOVCHAN

Department of Mathematical Sciences
Liverpool University

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 1999

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
(Data available)

ISBN 0 19 852770 5 (Hbk)
ISBN 0 19 852771 3 (Pbk)

Typeset using LaTeX

Printed in Great Britain
on acid-free paper by
T. J. International Ltd, Padstow, Cornwall

Preface to the revised edition

The revision of this book has been a source of both pleasure and pain to the authors. The pleasure has come from the opportunity to include new material, almost all of which is intended to unify and tie together further the existing topics, and to give the reader the best possible overview of the wonderful interplay between partial differential equations and their real-world applications. This is all in keeping with our unshakeable philosophy that partial differential equations offer fabulously effective data compression: the basically simple structure of many partial differential equations enables knowledge holders to make a quantitative model of almost any 'continuous' process going on around them.

The pain of revision has come from realising that on many occasions our zeal in writing the first edition overstretched our accuracy. However, we have made amends as scrupulously as we can; we have been immensely helped in this task, and with the incorporation of new material, by the helpful comments of many of our colleagues and collaborators. We are also very grateful to Alison Jones and colleagues at Oxford University Press for their invaluable assistance in the final stages of publication.

Oxford J. R. O.
S. D. H.
Edinburgh A. A. L.
Liverpool A. B. M.
January 2003

Preface to the first edition

In the 1960s, Alan Tayler, Leslie Fox and their colleagues in Oxford initiated 'Study Group' workshops in which academic mathematicians and industrial researchers worked together on problems of practical significance. They were soon able to show the world that mathematics can provide invaluable insight for researchers in many industries, and not just those which at the time employed professional mathematicians.

This message is the theme of Alan's book *Mathematical methods in applied mechanics* [43], which contains many examples of how mathematical modelling and applied analysis can be put to work. That book revealed the ubiquity of partial differential equation models, but it did not lay out a co-ordinated account of the theory of these equations from an applied perspective. Hence this complementary volume was planned in the 1980s, first emerging as very informal lecture notes.

Much has happened since then. Alan's illness brought about two authorship changes: first, Andrew Lacey and Sasha Movchan stepped in to help, and, after Alan's tragic death in 1995, Sam Howison became involved as well. Additionally, the past decade has seen many new practical illustrations and theoretical advances which have been incorporated into the book, while still keeping it at around first-year graduate level.

Only now can we see the debt we owe not only to Alan Tayler but also to those who have supported us over the past ten years. In particular, we thank June Tayler, Annabel Ralphs, Natasha Movchan and Hilary Ockendon for their forbearance, Brenda Willoughby for typing help at a crucial stage, and Elizabeth Johnston and her colleagues at Oxford University Press.

A book like this cannot be written without help from colleagues around the world, far too many to mention here, but we would especially like to acknowledge the many helpful comments we have received from post-docs, who are often the most important people at the interface between mathematics and the real world.

Oxford — J. R. O.
S. D. H.
Edinburgh — A. A. L.
Liverpool — A. B. M.
February 1999

Contents

Introduction

Partial differential equations are central to mathematics, whether pure or applied. They arise in mathematical models whose dependent variables vary continuously as functions of several independent variables, usually space and time. Their most striking attribute is their universality, a property which has enabled us to motivate every mathematical idea in this book by real-world examples from fluid or solid mechanics, electromagnetism, probability, finance and a host of other areas of application. Moreover, this applicability is growing day by day because of the flexibility and power of modern software tailored to suitable discretised approximations of the equations. Equally dramatic is the way in which the equations that arise in all these areas of application can so easily motivate the study of fundamental mathematical questions of great depth and significance and, conversely, benefit from the results of such investigations.

Whether or not it is in the context of a model of a physical situation, the analysis of a partial differential equation has many objectives. One of our principal goals will be to investigate the question of *well-posedness*. We will give a more precise definition of this in Chapter 2 but, roughly speaking, a partial differential equation problem is said to be well posed if it has a solution, that solution is unique, and it only changes by a small amount in response to small changes in the input data. The first two criteria are reasonable requirements of a sensible model of a physical situation, and the third is often expected on the basis of experimental observations. When thinking of well-posedness, we must also remember that it is often impossible to find explicit solutions to problems of practical interest, so that approximation schemes, and in particular numerical solutions, are of vital importance in practice. Hence, the question of well-posedness is intimately connected with the central question of scientific computation in partial differential equations: given the data for a problem with a certain accuracy, to what accuracy does the computed output of a numerical solution solve the problem? It is because the answer to this question is so important for modern quantitative science that well-posedness is a principal mathematical theme of this book.

Although many well-founded mathematical models of practical situations lead to well-posed problems, phenomena that are seemingly unpredictable, or at the least extremely sensitive to small perturbations, are not uncommon; examples include turbulent fluid flows described by the Navier–Stokes equations and dendrite growth modelled by the equations of solidification. Pure and applied mathematicians alike must therefore be prepared for both well-posed and ill-posed partial differential equation models. Chaos in scalar ordinary differential equations can occur if the order of the equation is at least three and so it is not surprising that what is effectively an infinite-order ordinary differential equation may have 'unpredictable' solutions. We must also remember that there are processes such as Brownian motion, which are random on a molecular scale, and yet have many properties which can be modelled by perfectly well-posed partial differential equations over much larger time and length scales. However, since we do not have the space to describe chaos theory, we will not be able to discuss the very interesting relationship between chaos and ill-posedness, although in Chapter 7 we will touch on several examples which have highly unpredictable behaviour. Nonetheless, we will be able to look at problems such as those involving exothermic chemical reactions where the model may be well-posed but its solution may only exist over a limited region in time and space before a singularity, or 'blow-up', occurs.

The advent of the computer has not only changed the attitude of the mathematical community to partial differential equations, but also the attitude of researchers in most fields where quantitative solutions of problems are now necessary. Powerful computers have encouraged people to attack so many hitherto intractable or novel problems that the number and variety of differential equations under study is increasing at an enormous rate. This observation brings us to the single most important practical reason for our writing this book, namely the 'data compression' implicit in a partial differential equation model. It is an astonishing fact that all the practical problems that we describe in this book, which range from paint flow to solidification, and from option pricing to combustion, can be described in a handful of symbols as the *quasilinear system*

$$\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \mathbf{u}}{\partial x_i} = \mathbf{b}, \tag{0.1}$$

together with suitable boundary conditions; here the unknown, $\mathbf{u}$, is a vector function of the independent variables x_i, $i = 1, \ldots, m$, while $\mathbf{A}_i$ and $\mathbf{b}$ are, respectively, square matrices and a vector which all depend on $\mathbf{u}$ and the x_i. It is the crucial fact that $\mathbf{A}_i$ and $\mathbf{b}$ do not depend on the derivatives of $\mathbf{u}$ that characterises quasilinearity. As we shall see later, we can even arrange for the right-hand side $\mathbf{b}$ to be $\mathbf{0}$.

To get some idea why this format is all-embracing, suppose we were confronted with a fairly general scalar first-order equation in two independent variables x, y in the form

$$\frac{\partial u}{\partial x} = G\left(x, y, u, \frac{\partial u}{\partial y}\right).$$

Setting $q = \partial u/\partial y$ and

$$\mathbf{u} = \begin{pmatrix} u \\ q \end{pmatrix},$$

after differentiating with respect to y, we find the system

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{\partial \mathbf{u}}{\partial x} + \begin{pmatrix} 0 & 0 \\ 0 & -\partial G/\partial q \end{pmatrix} \frac{\partial \mathbf{u}}{\partial y} = \begin{pmatrix} G \\ \partial G/\partial y + q\,\partial G/\partial u \end{pmatrix},$$

which is in quasilinear form.[1]

There is a dramatic difference between (0.1) and the ordinary differential equation when $m = 1$, namely

$$\mathbf{A}\frac{d\mathbf{u}}{dx} = \mathbf{b}.$$

In this latter case, as long as $\mathbf{A}$ is invertible, which it usually is, and $\mathbf{A}^{-1}\mathbf{b}$ satisfies an appropriate Lipschitz condition, there is a unique solution such that $\mathbf{u} = \mathbf{u}_0$ at some point $x = x_0$. However, it is clear that if $\mathbf{u} = \mathbf{u}(x, y)$ and

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} = \mathbf{b},$$

then, no matter how well behaved $\mathbf{A}$ and $\mathbf{b}$ are, we cannot solve this equation with $\mathbf{u}(x, y) = \mathbf{u}_0(x)$ at $y = y_0$ unless $\mathbf{A}\,\partial \mathbf{u}_0/\partial x = \mathbf{b}$.

This observation is the basis of our discussion in Chapter 1, which concerns the scalar case of (0.1) in which the term involving the highest derivative (which is called the *principal part* of the equation) is

$$\sum_{i=1}^{m} A_i \frac{\partial u}{\partial x_i}.$$

We will begin by identifying boundary data for which we might expect a solution to exist and data for which there is almost no hope of existence. This is the theme that pervades the subsequent two chapters, which deal with systems like (0.1) and simple scalar second-order equations, respectively. We will first have to worry about ill-posedness in Chapter 2; there we shall see that when $\mathbf{u}$ is given on some initial surface, we may well be able to find all its derivatives normal to that surface but that this information only enables us to continue $\mathbf{u}$ a very small distance away from the initial surface. However, it will become apparent in Chapters 3 and 5 that this restriction can sometimes be overcome by relaxing the requirement that all components of $\mathbf{u}$ be given on this surface.

In addition to cataloguing well-behaved and badly-behaved solutions for simple scalar second-order equations, Chapter 3 also provides an introduction to Chapters 4–6, each of which deals with a class of scalar second-order equations which occurs with unfailing regularity in branches of physics, engineering, chemistry, biology, and even social science and finance. Indeed, from the practical point of view of students wanting to know how to get an analytical feel for the solutions of equations falling into these classes, these chapters form the meat of the book and can be read more or less independently.

Chapter 7 is perhaps the most unusual one in the book because it addresses a class of problems that are rarely compiled outside the research literature. Yet recent

[1]Eagle-eyed readers will notice that the first matrix is eminently invertible (because the partial differential equation has been 'solved' for $\partial u/\partial x$), while the second is not (because information is lost when we differentiate). There is a lot more to this simple calculation, as we will see in §2.3 and §8.2. By the way, because we are aiming for a concise treatment, there are many footnotes in this book, so please do not be deterred by them; they mostly contain digressions from the main stream.

inroads of mathematical modelling into practical problems, especially those arising in industry, have revealed that many, many differential equation models have to be solved in regions that are unknown *a priori*. These regions must be found as part of the solution; typical examples are the melting of an ice cube or the sloshing of water in a container. We call these problems *free boundary problems* and, in Chapter 7, we endeavour to provide an entrée into the great body of knowledge that has grown up around them in recent years.

Despite the universality of (0.1), there are some advantages in studying fully nonlinear equations in their primitive form; in Chapter 8 we revert to problems in which $\mathbf{A}$ can indeed depend on $\partial u_i/\partial x_j$ as well as on $\mathbf{u}$. Thus (0.1) is no longer quasilinear and we will see that this means that we always encounter the possibility of non-existence or non-uniqueness when we attempt to find the derivatives of $\mathbf{u}$ in terms of its values on some known surface. This will be found to lead to many fascinating generalisations of the theory of non-quasilinear ordinary differential equations, such as envelope solutions and caustics, which means that geometric interpretations are even more valuable than in earlier chapters. Chapter 9 is a compendium of ideas concerning partial differential equations that do not fit conveniently into the earlier chapters: it could have gone on for ever.

One crucial mathematical idea that will emerge from the first six chapters is the value of being able to write down formally the solution of any *linear* partial differential equation, i.e. one in which $\mathbf{A}_i$ are independent of $\mathbf{u}$ and $\mathbf{b}$ is linearly dependent on $\mathbf{u}$ in (0.1). This idea is a generalisation of the one that says that, in order to solve a system of linear algebraic equations, we have to invert a matrix; instead of writing that $\mathcal{A}\mathbf{x} = \mathbf{b}$ usually implies $\mathbf{x} = \mathcal{A}^{-1}\mathbf{b}$, we say that $\mathcal{L}\mathbf{u} = \mathbf{f}$ usually implies $\mathbf{u} = \mathcal{L}^{-1}\mathbf{f}$. We will see that '$\mathcal{L}^{-1}$' can, when it exists, be expressed as an integral weighted by what is called a Green's function or Riemann function. However, finding this function or even some of its simple properties is almost always difficult and usually impossible. Hence readers should never be lulled into thinking that, because of their apparent conceptual simplicity, linear partial differential equations are either easy or boring.

There is one other remark we must make before we start. This is the regrettable fact that, in order to keep this book as short as it is, we have had to exclude almost all discussion of functional analysis, numerical methods, and in particular almost all discussion of the multitude of results that can be obtained by 'perturbation theory'. In fact, we will restrict attention to those results that can fairly easily be proved analytically or interpreted geometrically. It would have been easy in principle to double the length of most of the chapters by appending some of the important results that emerge from the relevant perturbation theory; it could have been doubled again had numerical methods been included, and yet again by describing the principal results from the modern function-analytic theory of partial differential equations. However, we emphasise that many of the results we obtain or cite would not have been discovered had not their originators experimented with approximate methods at the start.

Another advantage of our self-imposed restrictions is that the only prerequisites we hope the reader possesses are some familiarity with the idea of ordinary differential equations, functions of a single complex variable and the calculus of functions of several real variables. Most of all we would like them to know the Fredholm Alternative, but, in case this is unfamiliar, it is spelled out on p. 43. Although these are not demanding prerequisites, it will help if the reader can also bring to the book a relaxed mathematical attitude and a readiness to look at the broader picture: this is not a 'definition–theorem–proof' book, nor an exhaustive catalogue of methods and techniques. The authors' background is in physical applied mathematics, which inevitably slants some of the motivational examples and interpretations of the theory, but the basic message of well-posedness would have been the same had they been numerical analysts or probabilists. The fact that we have been able to eschew rigour and relegate certain calculations to the exercises means that we have been able to keep the book relatively short without compromising its applicability.

*To make a first reading easier, we have marked the harder sections of the text and exercises with an asterisk, so that they can be freely ignored by those who are pressed for time.

A bibliography, which consists almost entirely of related textbooks, is provided after Chapter 9.

1

First-order scalar quasilinear equations

1.1 Introduction

Even though this chapter deals only with the simplest category of partial differential equations, the theory that emerges is relevant to many important and fascinating practical situations. An example is the flow of a thin coat of paint down a wall, as illustrated in Fig. 1.1(a). Because the layer is thin, the velocity, say $u(x, y, t)$, is approximately unidirectional down the wall. Gravity is resisted by the viscosity of the paint, resulting in a shearing force, which we assume to be proportional to the velocity gradient $\partial u/\partial y$. A force balance on a small fluid element then shows that $\partial^2 u/\partial y^2$ is a constant, $-c$, which is proportional to gravity (see Fig. 1.1(b)). We assume that the paint sticks to the wall, so $u = 0$ on $y = 0$. Also, since the shearing force is zero on the paint surface $y = h(x, t)$, $\partial u/\partial y = 0$ there, and hence

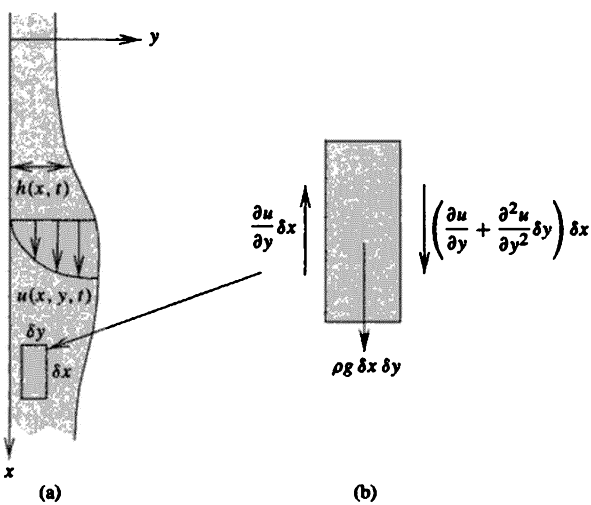


Fig. 1.1 (a) Paint on a wall. (b) Forces on a fluid element.

$$u = \frac{1}{2}cy(2h - y).$$

Finally, conservation of mass in the thin film gives that the time rate of change of the paint thickness must be balanced by the x-variation of the paint flow down the wall. This flux is $q(x,t) = \int_0^h u\,dy$; over a small time δt the amount lost from a small element of length δx is approximately $(q(x+\delta x,t) - q(x,t))\,\delta t$, and balancing this with the excess, $(h(x,t+\delta t) - h(x,t))\,\delta x$, gives

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}\int_0^h u\,dy = 0,$$

that is,

$$\frac{\partial h}{\partial t} + ch^2\frac{\partial h}{\partial x} = 0. \tag{1.1}$$

In fact, such first-order equations occur whenever we have a 'density', say $h(x,t)$, and a 'flux', say $q(h)$, which is a known function of h. Then conservation of mass implies the *kinematic wave equation*

$$\frac{\partial h}{\partial t} + \frac{\partial q(h)}{\partial x} = 0.$$

Equations of this kind are used to model, for example, the flow of traffic on a crowded road, blood flow through an elastic-walled tube such as an artery, or lubricant flowing down the mould wall of a continuous-casting machine; they also arise as special cases of the general theories of gas dynamics and hydraulics, as we shall see in Chapter 2.

In a completely different vein, suppose $p_n(t)$ is the probability that, after time t spent proof-reading an arduous text on differential equations, the draft still contains n errors. Let us also assume that in a short time interval $(t, t+\delta t)$ the authors find and correct one and only one error with probability $\mu n p_n(t)\delta t$. Since a draft with n errors at time $t+\delta t$ can only result from ones in which there were n or $n+1$ errors at time t, a simple counting argument using conditional probabilities gives

$$p_n(t+\delta t) = \mu(n+1)p_{n+1}(t)\delta t + (1 - \mu n\,\delta t)p_n(t).$$

Thus, as $\delta t \to 0$, we find that the generating function $p(x,t) = \sum_{n=0}^{\infty} p_n(t)x^n$ satisfies

$$\sum_{n=0}^{\infty}\frac{dp_n}{dt}x^n = \sum_{n=0}^{\infty}(\mu(n+1)p_{n+1} - \mu n p_n)x^n,$$

and hence that

$$\frac{\partial p}{\partial t} + \mu(x-1)\frac{\partial p}{\partial x} = 0.$$

Both this equation and (1.1) must be supplemented by suitable initial or boundary conditions. Thus, we might prescribe the initial paint thickness $h(x,0)$, while a sensible condition for the proof-reading model would be to assume that there are N errors when proof-reading starts, so that $p(x,0) = x^N$. Note, though, that we might also want to specify other conditions for the paint model; for example, if paint is applied by a brush moving with speed V, we might give h at $x = Vt$.

1.2 Cauchy data

Motivated by these examples and, obviously, by the principle that it is always best to start from the simplest situation, we now develop a theory for general first-order *quasilinear* equations. These are defined to be equations of the form

$$a(x,y,u)\frac{\partial u}{\partial x}+b(x,y,u)\frac{\partial u}{\partial y}=c(x,y,u), \tag{1.2}$$

where it is important that a, b and c do not depend on the derivatives of u; we also assume that a, b and c are smooth (i.e. continuously differentiable) functions of their arguments.

Suppose that Γ is a curve in the (x,y) plane; we define *Cauchy data* to be the prescription of u on Γ. It is convenient to write this boundary condition in the parametric form

$$x=x_0(s),\quad y=y_0(s),\quad u=u_0(s)\quad \text{for } s_1\leqslant s\leqslant s_2. \tag{1.3}$$

Here x_0, y_0 and u_0 are smooth[2] functions of s such that there is no value of s for which $x_0'(s)=y_0'(s)=0$, the prime denoting differentiation with respect to s.

The boundary value problem of finding a continuously differentiable function u satisfying (1.2) and (1.3) may be interpreted geometrically as that of constructing a surface $u=u(x,y)$, in the space (x,y,u), called the *integral* or *solution surface*, which satisfies (1.2) and passes through the boundary curve defined by (1.3). This is an obvious extension of the initial value problem for an ordinary differential equation, which requires the construction of a curve (i.e. a graph) passing through a boundary point. For an ordinary differential equation, say $a(x,u)\,\partial u/\partial x=c(x,u)$, this is generally possible unless $a=0$ at the boundary point, and for the partial differential equation a subset of boundary curves is unacceptable for a similar reason.

If we differentiate the boundary data with respect to s along the boundary curve, then we find

$$u_0'=\frac{\partial u}{\partial x}x_0'+\frac{\partial u}{\partial y}y_0'. \tag{1.4}$$

Equations (1.4) and (1.2) evaluated on Γ, where a, b and c are known functions of s, are two simultaneous equations for $\partial u/\partial x$ and $\partial u/\partial y$ on Γ. Hence $\partial u/\partial x$ and $\partial u/\partial y$ are determined uniquely on Γ, provided that the determinant of coefficients

$$\begin{vmatrix} a & b \\ x_0' & y_0' \end{vmatrix}=ay_0'-bx_0'\neq 0\quad\text{for all } s_1\leqslant s\leqslant s_2. \tag{1.5}$$

If this condition is satisfied and if a, b, c and the Cauchy data have Taylor series at each point, it can easily be verified that the partial derivatives of u of all orders are defined uniquely by the Cauchy data and the partial differential equation, and a Taylor series expansion for $u(x,y)$ about a point on Γ can be constructed

[2] When x_0, y_0 or u_0, or a, b, c are not smooth the problem needs a completely new approach, as we shall see in §1.6.

formally. This forms the starting point for the Cauchy–Kowalevski existence and uniqueness theorem, which considers the convergence of this Taylor series and guarantees a unique solution if condition (1.5) is satisfied; it will be discussed further in Chapter 2.

If condition (1.5) is *not* satisfied at a point $s = s_0$ (or for a range of values of s) then, in general, no solution exists in the neighbourhood of the boundary curve. However, in the special case when

$$\frac{x_0'}{a} = \frac{y_0'}{b} = \frac{u_0'}{c}, \tag{1.6}$$

(1.2) and (1.4) are linearly dependent and $\partial u/\partial x$ and $\partial u/\partial y$ exist but are not uniquely determined at $s = s_0$. In this case many solution surfaces may pass through a single boundary curve, and this suggests that curves satisfying these ordinary differential equations have an important role to play, as we shall see in the next section.

1.3 Characteristics

The partial differential equation (1.2),

$$a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} = c,$$

has a geometrical interpretation which is also the key to its solution given Cauchy boundary data $x = x_0(s)$, $y = y_0(s)$ and $u = u_0(s)$ satisfying (1.5). The normal to the solution surface in (x, y, u) space is $(\partial u/\partial x, \partial u/\partial y, -1)$, and we can rewrite the partial differential equation in the form

$$(a, b, c) \cdot \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, -1\right) = 0.$$

Thus, (a, b, c) lies in the tangent plane to the solution surface at each point. Hence, if we construct a curve $(x(t), y(t), u(t))$, parametrised by t, by solving the ordinary differential equations

$$\begin{aligned} \dot{x} &= \frac{\mathrm{d}x}{\mathrm{d}t} = a(x, y, u), \\ \dot{y} &= \frac{\mathrm{d}y}{\mathrm{d}t} = b(x, y, u), \\ \dot{u} &= \frac{\mathrm{d}u}{\mathrm{d}t} = c(x, y, u), \end{aligned} \tag{1.7}$$

then this curve lies in a solution surface for all t. If, in addition, we require that at $t = 0$

$$x = x_0(s), \quad y = y_0(s), \quad u = u_0(s), \tag{1.8}$$

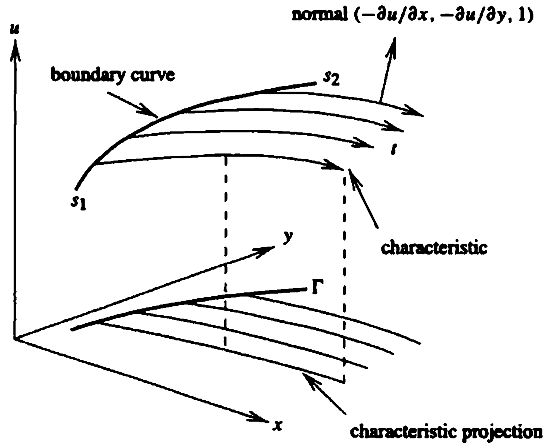


Fig. 1.2 Generation of the solution surface by characteristics.

then this solution surface also passes through the boundary curve. The solution curves of (1.7) are called the *characteristics*;[3] as s varies, the family of characteristics generates a surface, as in Fig. 1.2, which we hope is the required solution surface. As also indicated in Fig. 1.2, the projections of the solution curves of (1.7) onto the (x, y) plane are called the *characteristic projections*. It is not, however, clear from this argument that the surface so constructed is smooth, or even continuous; there might, for example, be kinks at which the derivatives of u are discontinuous, or the surface might turn over on itself.

If the solution of the partial differential equation $a\,\partial u/\partial x + b\,\partial u/\partial y = c$ is required to be what is called a *classical solution*, then u and its first derivatives must exist and make the left- and right-hand sides equal at each point. Despite the plausibility of our geometric picture, the following analytical argument is necessary to verify this.

Since a, b and c are functions of x, y and u but not t explicitly, the characteristic equations (1.7) are a system of three autonomous ordinary differential equations subject to the initial data (1.8). Now a, b and c are Lipschitz continuous, so the Cauchy–Picard theorem implies that the characteristic equations have a unique local solution. Hence we may write

$$x = x(s,t), \quad y = y(s,t), \quad u = u(s,t), \tag{1.9}$$

where $x(s,0) = x_0(s)$, $y(s,0) = y_0(s)$ and $u(s,0) = u_0(s)$. Moreover, from the same theorem, x, y and u are differentiable functions of both s and t; they also depend continuously on the boundary data.

[3]We will shortly give two other important reasons for considering the curves defined by (1.7), which will motivate new definitions of characteristics applicable to more general partial differential equations.

The relationships $x = x(s,t)$, $y = y(s,t)$ and $u = u(s,t)$ are a parametric representation of the solution surface, which gives u locally as a unique differentiable function of x and y if there is a unique inverse of the transformation $(s,t) \mapsto (x,y)$. From the inverse function theorem, the necessary and sufficient condition for this is that

$$\begin{vmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \end{vmatrix} = \begin{vmatrix} x' & \dot{x} \\ y' & \dot{y} \end{vmatrix} \neq 0, \infty,$$

where, here and henceforth, we write $' = \partial/\partial s$, $\dot{} = \partial/\partial t$. Using (1.7) this reduces to

$$ay' - bx' \neq 0, \infty,$$

which is certainly satisfied on the boundary curve $t = 0$, where it reduces to the condition (1.5); by continuity, it is also satisfied in some neighbourhood of the boundary. We thus have a local existence result for u. Moreover, it is trivial to show that the function so constructed satisfies the original differential equation $a\,\partial u/\partial x + b\,\partial u/\partial y = c$; we simply differentiate along the characteristics to give

$$\dot{u} = c = \frac{\partial u}{\partial x}\dot{x} + \frac{\partial u}{\partial y}\dot{y} = a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y}$$

from (1.7). Note that this result shows that along a characteristic the partial derivative terms in (1.2) reduce to a directional derivative of u in that direction.

1.3.1 Linear and semilinear equations

Although we have just derived a local existence result, there is still, of course, the possibility that u develops singularities further away from Γ, and we will soon see that this can easily happen for equations that are not linear. Linear equations are those in which $c(x,y,u) = \alpha(x,y)u + \beta(x,y)$ and a, b are independent of u. The simplest example is that of kinematic waves with $\alpha = \beta = 0$ and a, b both constant. Then the characteristic equations are

$$\dot{x} = a, \quad \dot{y} = b, \quad \dot{u} = 0,$$

so that

$$x = at + x_0(s), \quad y = bt + y_0(s), \quad u = u_0(s),$$

which is another way of saying $u = f(bx - ay)$ for some arbitrary function f; the boundary values simply propagate along the characteristic projections.

We now turn to a wider class of equations for which the characteristic equations (1.7) can still be used easily. These are *semilinear* equations, defined to be those that are linear in their principal part, namely the terms involving the highest derivatives. Thus, for $a\,\partial u/\partial x + b\,\partial u/\partial y = c$, these equations are defined to be such that the left-hand side, which contains all the derivatives, is linear in u in that a and b depend on x and y alone; however, c may now depend nonlinearly on the dependent variable u.

The characteristic equations (1.7) now reduce to a first-order differential equation since the equation for u uncouples and

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot{y}}{\dot{x}} = \frac{b}{a}.$$

This equation has integral curves in the (x, y) phase-plane, which are non-intersecting except possibly at a *critical point* where $a = b = 0$.[4] Subject to the initial conditions $x(s, 0) = x_0(s)$ and $y(s, 0) = y_0(s)$, they constitute a global one-parameter family of curves C in the (x, y) plane, namely the characteristic projections (often, these curves C are themselves called the characteristics). There is a unique characteristic projection, parametrised by t and labelled by s, through any non-critical point; hence the map $(s, t) \mapsto (x, y)$ is invertible there, and so $ay' - bx' \neq 0$.

Assuming the characteristic projections are known, we can calculate how u varies along them from (1.7), using either

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{c}{a} \quad \text{or} \quad \frac{\mathrm{d}u}{\mathrm{d}y} = \frac{c}{b},$$

whichever is more convenient. Of course, if c depends nonlinearly on u, these last ordinary differential equations might not have global solutions, as we will see in Example 1.3, but the fact that (1.7) has decomposed into two first-order scalar ordinary differential equations, which can be solved sequentially, is a very substantial simplification, even if an explicit solution cannot be found in terms of elementary functions. The following example illustrates what can happen when we are lucky.

Example 1.1 Solve the following linear problem:

$$x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = (x + y)u \qquad \text{with} \qquad u = 1 \quad \text{on } x = 1,\ 1 < y < 2;$$

the initial data can be written in the parametric form

$$x_0(s) = 1, \quad y_0(s) = s, \quad u_0(s) = 1 \quad \text{for } 1 < s < 2.$$

In this example we expect singular behaviour to occur at $x = y = 0$, by analogy with the theory of ordinary differential equations. The characteristic equations are

$$\dot{x} = x, \quad \dot{y} = y, \quad \dot{u} = (x + y)u,$$

and the characteristic projections C are given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{y}{x}.$$

As shown in Fig. 1.3, they are straight lines through the origin, which is a critical point.

[4]See Exercise 1.9 for an example of behaviour near a critical point.

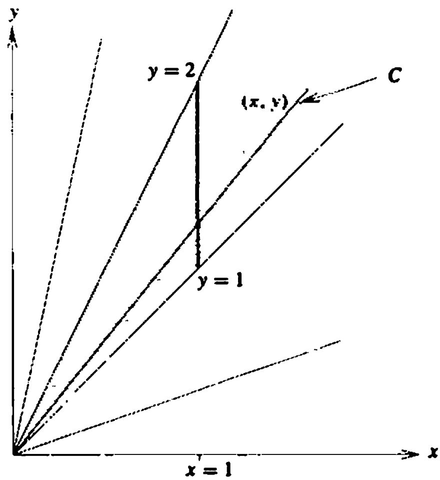


Fig. 1.3 Characteristic projections for Example 1.1.

Integrating $\mathrm{d}u/\mathrm{d}x = (x+y)u/x$ along a characteristic C, on which y/x is constant, and setting $u = 1$ at $x = 1$, we have

$$\log u = \int_C \left(1 + \frac{y}{x}\right) \mathrm{d}x = \left(1 + \frac{y}{x}\right)(x - 1).$$

This solution for u is only defined in the shaded wedge in Fig. 1.3; moreover, u is singular as $x \to 0+$, a result which is also apparent if we calculate the solution in parametric form:

$$x = \mathrm{e}^t, \quad y = s\mathrm{e}^t, \quad \log u = (1+s)\int_0^t \mathrm{e}^\tau \, \mathrm{d}\tau = (1+s)(\mathrm{e}^t - 1).$$

With $x > 0$, the Jacobian $|\partial(x,y)/\partial(s,t)| = -\mathrm{e}^{2t}$ is never zero, so that s and t may be eliminated locally to obtain u as a unique function of x and y. As $t \to -\infty$, $(x,y) \to (0,0)$, the Jacobian tends to zero, and u is singular. We will often use the jargon ***blow-up*** as a shorthand description for singularity development in either u or its derivatives.

1.4 Domain of definition and blow-up

In the example above, the region of $x > 0$, $y > 0$ in which the solution exists is enclosed by the characteristic projections through the end-points of the curve in the (x, y) plane on which the boundary data is given, i.e. the projection of the boundary curve. This region is known as the ***domain of definition***. If the boundary condition had instead been $u = 1$, $x = 1$ for $y_1 \leqslant y \leqslant y_2$, then the domain of definition of the solution would have been $y_1 \leqslant y/x \leqslant y_2$, $x > 0$. A general property for all quasilinear scalar equations is that the domain of definition is, at the very least, limited by the characteristic projections through the end-points of the projection of the boundary curve. Further limitations on the domain of definition may result if a and b vanish, or if u blows up as we integrate along a characteristic, or if the Jacobian $|\partial(x,y)/\partial(s,t)|$ vanishes on some curve in the (x,y) plane. In fact, the domain of definition of the solution could never extend beyond such curves since, on them, at least one of the derivatives of u is usually unbounded.

Note that the Jacobian vanishes where $ay' = bx'$ but, in the general quasilinear case, this limitation on the domain of definition cannot be determined without first finding the solution u. For the semilinear case, however, $\partial u/\partial x$ and $\partial u/\partial y$ are uniquely defined except possibly at critical points where $a = b = 0$, and, as remarked earlier, the Jacobian $|\partial(x,y)/\partial(s,t)|$ is necessarily non-zero except at a critical point.

An example where the Jacobian vanishes on the boundary on which Cauchy data is prescribed is the following.

Example 1.2 Solve

$$x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = (x+y)u, \tag{1.10}$$

(the same equation as in the previous example) with $u = 1$ on the circular arc Γ defined by

$$x_0 = 2 - \sqrt{2}\cos s, \quad y_0 = \sqrt{2}\sin s \quad \text{for } 0 \leqslant s < \tfrac{\pi}{4}.$$

The limiting characteristic projections are $y = 0$ and $y = x$, and the domain of definition is therefore $0 < y < x$, as shown in Fig. 1.4.

However, there is a worry here because, when we evaluate the solution parametrically, we find

$$x = \left(2 - \sqrt{2}\cos s\right)\mathrm{e}^t, \quad y = \sqrt{2}\sin s\,\mathrm{e}^t,$$
$$\log u = \left(2 - \sqrt{2}\cos s + \sqrt{2}\sin s\right)\left(\mathrm{e}^t - 1\right),$$

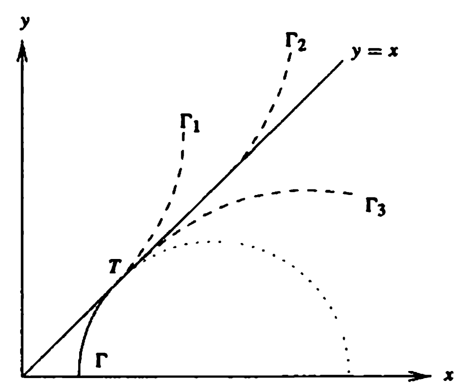


Fig. 1.4 Example 1.2: the solution is only defined in $0 < y < x$.

and hence

$$x'\dot{y} - y'\dot{x} = 2e^{2t}\left(1 - \sqrt{2}\cos s\right);$$

although this expression is non-zero for $0 \leqslant s < \pi/4$, it vanishes where Γ is tangent to the characteristic $s = \pi/4$. This sends us a danger signal, not for the problem (1.10) as it stands, but for any problem in which we might attempt to continue the Cauchy data up to or beyond the point T in Fig. 1.4. If we were to continue on Γ_1, we might only have problems on the characteristic $y = x$, but if we tried Γ_2 (any curve containing a segment of the characteristic $y = x$) or Γ_3, we would almost certainly pose ourselves a problem with no solution at all. In the case of Γ_2, our data would probably violate the characteristic equations (1.7), and for Γ_3, which meets the characteristic projections passing through Γ, the data would probably be inconsistent with the information propagating along the characteristics from Γ.

It is well known that the solutions to first-order nonlinear ordinary differential equations can easily blow-up by developing singularities. The same can therefore happen for semilinear equations, as the following example shows.

Example 1.3 Solve

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = u^2 \quad \text{in } x > 0 \tag{1.11}$$

with

$$u = e^{-y^2} \quad \text{on } x = 0, \qquad \text{i.e.} \qquad x_0 = 0, \quad y_0 = s, \quad u_0 = e^{-s^2}.$$

Clearly, the parametric form of the solution is

$$x = t, \quad y = t + s, \quad u = \frac{1}{e^{s^2} - t}.$$

The nonlinear right-hand side of (1.11) is so strong that the solution blows up, with $u \to \infty$, on the curve $x = e^{(y-x)^2}$. The domain of definition is the region where $x < e^{(y-x)^2}$; there is no obvious way to define u beyond this region without some drastic modification to our partial differential equation.

This kind of behaviour is rarely encountered in practice for first-order scalar equations, but we will see many examples of it for higher-order equations in Chapter 6. However, there is another more endemic kind of blow-up that occurs for quasilinear first-order equations, to which we now turn.

1.5 Quasilinear equations

The general theory of integrals of three autonomous ordinary differential equations $\dot{x} = a$, $\dot{y} = b$ and $\dot{u} = c$ is much more difficult than that for the two such equations which were obtained in the linear or semilinear case. Indeed, the extra degree of freedom enjoyed by three-dimensional phase-spaces compared to phase-planes means that we must be prepared for chaotic characteristics and hence 'wrinkled' integral surfaces, if indeed the integral surfaces can be defined at all. We therefore begin by looking at quasilinear partial differential equations by example.

When it comes to writing down explicit solutions, we may be fortunate enough to encounter an equation as symmetric as

$$x(y-u)\frac{\partial u}{\partial x}+y(u-x)\frac{\partial u}{\partial y}=u(x-y), \tag{1.12}$$

in which case it is easy to spot integrals of

$$\frac{\mathrm{d}x}{x(y-u)}=\frac{\mathrm{d}y}{y(u-x)}=\frac{\mathrm{d}u}{u(x-y)}$$

in the form[5]

$$x+y+u=x_0(s)+y_0(s)+u_0(s), \quad xyu=x_0(s)y_0(s)u_0(s), \tag{1.13}$$

so that the characteristic projections are given by

$$x+y+\frac{x_0(s)y_0(s)u_0(s)}{xy}=x_0(s)+y_0(s)+u_0(s).$$

From the practical viewpoint, there is another serendipitous approach to the characteristic equations which is often more convenient than working with the parametric representation in terms of s and t. This is to note that, if we can spot two independent integrals $f(x,y,u)$ = constant and $g(x,y,u)$ = constant, then the general solution of the original partial differential equation (1.2) is $G(f,g)=0$, where G is arbitrary. To see this, note that the surface $G=0$ meets a surface $f=c$ where $G(c,g)=0$, i.e. where g = constant; hence the surface $G=0$ is composed of characteristics in accordance with Fig. 1.2, and this reflects the fact that any two integral surfaces intersect in a characteristic. Clearly, the solution to the example above can be obtained in this way, with $f=xyu$ and $g=x+y+u$; see Exercise 1.4. However, at the end of the day, the existence of explicit solutions is a rare event indicative of some underlying symmetry which may be difficult to discern; we will return to this general question in Chapter 6.

Parametric solution representations such as (1.13) may conceal a phenomenon which can be studied conveniently by the simpler prototype

$$\frac{\partial u}{\partial x}+u\frac{\partial u}{\partial y}=0. \tag{1.14}$$

This is a kinematic wave in which x plays the role of time; the wave speed depends linearly on the wave amplitude u, rather than quadratically as in the paint model (1.1). We can get a good idea of what might happen here by watching waves run up a shallow beach, where water at the crest of a wave moves faster than at

[5]In this context the *componendo et dividendo* rule

$$\frac{a_1}{b_1}=\frac{a_2}{b_2}=\cdots=\frac{a_n}{b_n}=\lambda \quad \Rightarrow \quad \frac{\sum_i \lambda_i a_i}{\sum_i \lambda_i b_i}=\lambda,$$

for any λ_i, is often useful.

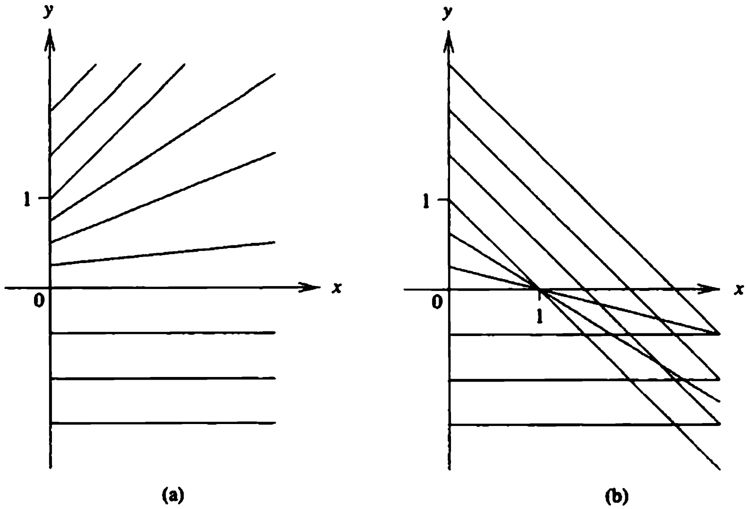


Fig. 1.5 Characteristic projections of the solution of (1.14) and (1.15): (a) positive ramp; (b) negative ramp.

a trough; in fact, solutions u of (1.14) can be shown to be a crude model of the elevation of such waves. We consider Cauchy data appropriate to an initial value problem in which $u = f(y)$ on $x = 0$, so that

$$x_0 = 0, \quad y_0 = s, \quad u_0 = f(s),$$

and the solution is

$$x = t, \quad y = tf(s) + s, \quad u = f(s).$$

The general integral is given implicitly by $u = f(y - ux)$ for any f.

We have the very helpful result that u is constant on the characteristic projections $\mathrm{d}y/\mathrm{d}x = u$, which are thus straight lines. Now consider the implications for two kinds of 'ramp' data[6] in which

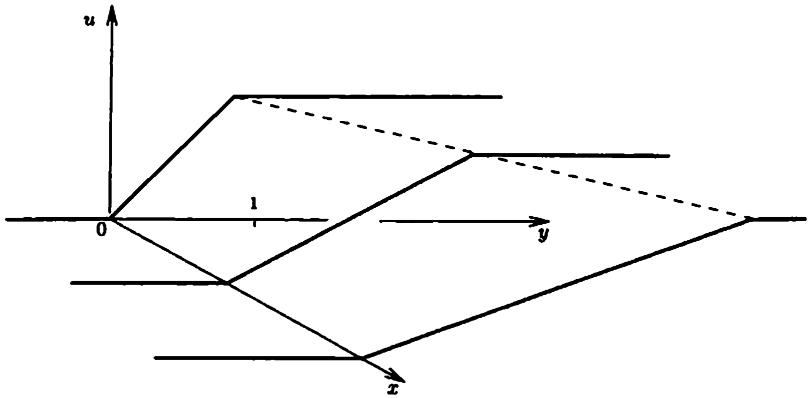
$$f_{\pm}(y) = \begin{cases} 0, & y < 0, \\ \pm y, & 0 \leqslant y < 1, \\ \pm 1, & 1 \leqslant y. \end{cases} \tag{1.15}$$

The solution for $x > 0$ is given by considering the characteristic projections, shown in Fig. 1.5, as

$$u_{\pm} = \begin{cases} 0, & y < 0, \\ y/(x \pm 1), & 0 \leqslant y < 1 \pm x, \\ \pm 1, & 1 \pm x \leqslant y. \end{cases}$$

[6]Such a Cauchy problem violates our crucial earlier assumption of only considering smooth data, but everything we are about to say would apply if we were to go through the chore of replacing (1.15) by a smooth approximation and then taking a suitable limit.

(a)



(b)

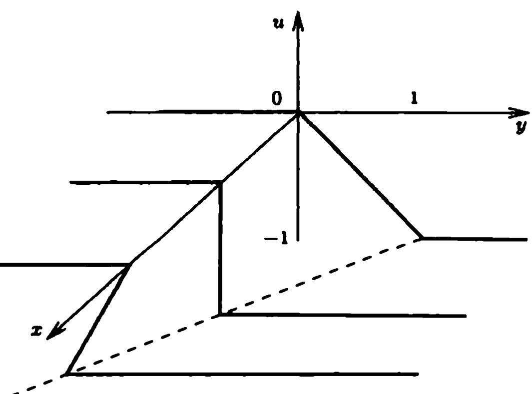


Fig. 1.6 Solution of (1.14) and (1.15): (a) positive ramp; (b) negative ramp.

The initial discontinuity in slope at $y = 0$ remains fixed on the characteristic projection $y = 0$, but the slope discontinuity at $y = 1$ is propagated along the characteristic projection $y = 1 \pm x$, and the profile of $u_\pm$ as a function of y remains piecewise linear, as in Fig. 1.6. Thus the ramp u_+ of positive slope becomes shallower as x increases, whereas the ramp u_- of negative slope steepens until, at $x = 1$, it is vertical; for $x > 1$, the profile is triple-valued. For the ramp of positive slope, there is no value of $x > 0$ for which the solvability condition $\partial(x, y)/\partial(s, t) \neq 0, \infty$ is violated but, for the negative slope ramp, there is a curve in $x \geqslant 1$ on which this Jacobian vanishes. We emphasise that this statement applies even if we 'smooth' the ramp because it is an inevitable consequence of the nonlinearity in (1.14). This 'turning-over' phenomenon is familiar to anyone who has applied paint too thickly to a vertical wall; the model (1.1) is susceptible to a simple modification of the analysis just given. The same remark applies to traffic pile-ups, or other situations described by the kinematic waves that we mentioned at the beginning of the chapter. However, on the mathematical side, the phenomenon has serious implications for the whole study of nonlinear partial differential equations because these kinds of discontinuities are likely to be encountered whenever we try to use nonlinear equations as mathematical models.

Before we consider what to do about this, let us also note that there is another aspect to partial differential equation models involving discontinuities. This is the observation that many physical phenomena that we would like to model can only be represented at all conveniently by differential equations, possibly even linear ones, whose solutions have discontinuities across boundaries which themselves have to be determined as a part of the solution. Such configurations are called *free boundary problems* and we will devote a chapter to them later on. For the moment we note that they could occur when modelling heat flow in a material that can change phase or otherwise react chemically in a very thin layer, such as a flame; in either case there is a discontinuity in the heat flux at the free boundary. Equally, a free surface between two immiscible fluids can be considered as a discontinuity in density, or a tumour growth boundary as a discontinuity in the concentration of some biological agent. The only difference in principle between these examples and (1.14) is that, in the latter case, the discontinuity occurs spontaneously as x increases, while the above-mentioned physical examples have their singularity imposed in the prescription of the problem.

This state of affairs motivates us to consider a more imaginative approach to partial differential equations than that of prosaically searching for what we have called classical solutions, in which the dependent variable and its derivatives are all smooth enough to satisfy the differential equation *everywhere*. We could be ambitious, and allow solutions to be singular at points (which could be relevant to, say, models for explosions) or on lines (to model, say, vortices in fluids or superconductors), but here we will only consider discontinuities across surfaces of one dimension fewer than the dimension of the space of independent variables (i.e. curves, as far as most of this chapter is concerned).

1.6 Solutions with discontinuities

Our first naive idea based on (1.14) (and its yet-to-be-revealed relevance to, say, shock waves in gas dynamics) is to contemplate the idea of a 'weak', as distinct from 'classical', solution as a combination of classical solutions, each defined on a different domain. These domains are then going to be patched together in such a way that, across the boundaries between domains on which there are discontinuities in some derivatives, a suitably generalised form of the governing equation (1.2),

$$a\frac{\partial u}{\partial x}+b\frac{\partial u}{\partial y}=c,$$

is satisfied. The definition will be made precise, even when u itself is discontinuous, in the next section, but let us consider first the problem of looking for a solution for which only the first derivatives of u may be discontinuous across some curve

C in the (x, y) plane, while u itself is continuous there. If the curve is defined by $x = x(t)$, $y = y(t)$ and we denote the limiting values of functions as this curve is approached from either side by superscripts $^+$ and $^-$, then

$$\dot{u}^+ = \dot{x}\frac{\partial u}{\partial x}^+ + \dot{y}\frac{\partial u}{\partial y}^+, \quad \dot{u}^- = \dot{x}\frac{\partial u}{\partial x}^- + \dot{y}\frac{\partial u}{\partial y}^- .$$

Since u is continuous across C, so is $\dot{u}$, and therefore

$$\dot{x}\left[\frac{\partial u}{\partial x}\right]_-^+ + \dot{y}\left[\frac{\partial u}{\partial y}\right]_-^+ = 0, \tag{1.16}$$

where the square bracket notation denotes the 'jump' or finite discontinuity in the expression across C. The function u must also satisfy the partial differential equation except on C, so by subtraction we have

$$a\left[\frac{\partial u}{\partial x}\right]_-^+ + b\left[\frac{\partial u}{\partial y}\right]_-^+ = 0, \tag{1.17}$$

where a and b are evaluated on C without ambiguity since u is continuous there. The necessary condition for these gradient discontinuities to be non-zero is $b\dot{x} = a\dot{y}$, and therefore C must be a characteristic projection.

Thus a further interpretation of characteristics is that they are lines on the solution surface across which there need not be a continuously turning tangent plane, but instead there is a 'kink' or an 'edge'.[7] Such a solution is generated when the boundary curve only has a piecewise continuous tangent vector; it is not strictly speaking a classical solution since the partial differential equation is not satisfied on C. However, C *does* separate regions in which classical solutions *can* be defined and calculated without having to consider the size of the discontinuity.

In the linear case, with $c(x, y, u) = \alpha(x, y)u + \beta(x, y)$, we can compute the magnitude of the discontinuity as it *propagates* along C as follows. First observe that, if $\dot{x} = a$ and $\dot{y} = b$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial u}{\partial x}\right]_-^+ = a\left[\frac{\partial^2 u}{\partial x^2}\right]_-^+ + b\left[\frac{\partial^2 u}{\partial x \partial y}\right]_-^+ .$$

Differentiating the original partial differential equation with respect to x and subtracting values on either side of C, we obtain

$$a\left[\frac{\partial^2 u}{\partial x^2}\right]_-^+ + b\left[\frac{\partial^2 u}{\partial x \partial y}\right]_-^+ + \frac{\partial a}{\partial x}\left[\frac{\partial u}{\partial x}\right]_-^+ + \frac{\partial b}{\partial x}\left[\frac{\partial u}{\partial y}\right]_-^+ = \alpha\left[\frac{\partial u}{\partial x}\right]_-^+ .$$

Hence from (1.17), and assuming $b \neq 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial u}{\partial x}\right]_-^+ = \left(\alpha - \frac{\partial a}{\partial x} + \frac{a}{b}\frac{\partial b}{\partial x}\right)\left[\frac{\partial u}{\partial x}\right]_-^+, \tag{1.18}$$

so that $[\partial u/\partial x]_-^+$ never vanishes if it is non-zero at $t = 0$.

[7] Indeed, this interpretation motivates an alternative starting point for our definition of characteristics which, unlike that in §1.3, can be generalised to vector partial differential equations, as we will see in Chapter 2.

Motivated by the discussion in §1.5, we now consider a situation in which u itself may be discontinuous. The dangers inherent in doing this can be illustrated by looking at a class of problems of which the following is a simple example. Consider the Cauchy problem

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0, \quad u(0, y) = \begin{cases} 0, & y < 0, \\ 1, & y > 0, \end{cases}$$

and let us enquire whether there is any special curve $y = f(x)$, $f(0) = 0$, such that the solution is

$$u = \begin{cases} 0, & y < f(x), \\ 1, & y > f(x). \end{cases} \tag{1.19}$$

We can give a plausible answer even to the corresponding question for the general semilinear case, in which the boundary data has a discontinuity in u at the point $s = s_0$, where $s_1 < s_0 < s_2$. We know that there is a unique classical solution in the domain bounded by the projections of the characteristics through the points defined by s_1 and $s_0 - 0$, and a further classical solution in the domain bounded by the characteristic projections through $s_0 + 0$ and s_2. In the semilinear case the characteristic projection is determined independently of the value of u, so that a unique characteristic projection C_0 through $s = s_0$ is defined (Fig. 1.7). Thus a classical solution is defined in the two domains $D_\pm$ which are adjacent and separated by C_0. Hence one way of defining a generalised solution in a unique way is to permit u to have a jump, or finite discontinuity, only across a characteristic projection such as C_0. This implies that $f(x) = x$ in (1.19).

The size of the discontinuity in u is given by the limiting values of u on either side of C_0, which have to satisfy the characteristic equations (1.7), so that

$$\frac{\mathrm{d}}{\mathrm{d}t}[u]_-^+ = [c]_-^+ .$$

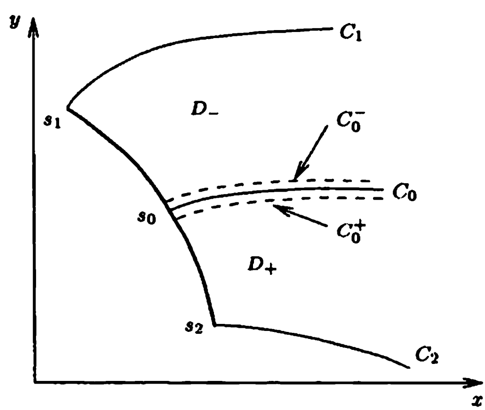


Fig. 1.7 Discontinuous boundary data for a semilinear equation.

If c is independent of u, the discontinuity in u remains constant; if $c = \alpha(x,y)u$, then, on C_0,

$$[u]_-^+ = A \exp \int_0^t \alpha\big(x(\tau), y(\tau)\big)\, d\tau,$$

where A is constant.

However, in the quasilinear case the limiting characteristic projections C_0^+ and C_0^- are in general different since they depend on the values u_0^+ and u_0^-, respectively. Thus the two domains in which the classical solutions are defined may overlap or not be adjacent, leading to multi-valuedness or non-existence, respectively. In this case the existence of a discontinuity in the boundary data alters the solution over a region larger than just the one characteristic. A new approach is needed, and it will be described in the next section.

Before we proceed, we emphasise that we cannot be too careful in constructing any theory of discontinuous solutions of partial differential equations. The moment we allow the solution to be ill-behaved in any way that is not natural for the equation, we may be like a fisherman who has cast his net so widely as to catch every fish in the river. If this happens to us, we must be prepared to have to filter out many candidate solutions which should not have been allowed in the first place.

*1.7 Weak solutions

On a curve C_0 across which u or its first derivatives have discontinuities, the partial differential equation $a\,\partial u/\partial x + b\,\partial u/\partial y = c$ has no meaning. However, we have seen in the previous section that jumps, i.e. finite discontinuities, in the first derivatives of u can be described by piecing together different classical solutions across characteristics systematically and naturally. The motivation for this section is to find a similarly systematic way of describing finite discontinuities in u itself. Such discontinuities are often called *shocks*, a term originating from partial differential equation models in gas dynamics. When shocks are present, the problem is underdetermined when stated just as a partial differential equation; either a condition on the value of the discontinuity must be given, or the problem must be rewritten in such a way that it is meaningful even in the presence of shocks.

For ease of exposition we begin by illustrating the latter procedure for the linear equation

$$a(x,y)\frac{\partial u}{\partial x} + b(x,y)\frac{\partial u}{\partial y} = \alpha(x,y)u. \tag{1.20}$$

We introduce an arbitrary differentiable function $\psi(x,y)$, called a *test function*, so that if u is differentiable and satisfies (1.20) then, for all ψ,

$$\frac{\partial}{\partial x}(au\psi) + \frac{\partial}{\partial y}(b\psi u) \equiv u\left(\frac{\partial}{\partial x}(a\psi) + \frac{\partial}{\partial y}(b\psi) + \alpha\psi\right). \tag{1.21}$$

Thus, if u is prescribed on some initial arc Γ and we integrate over a region D between Γ and an arbitrary curve γ, and ψ is restricted to vanish on γ (see Fig. 1.8(a)), then (1.21) allows us to use Green's theorem to obtain

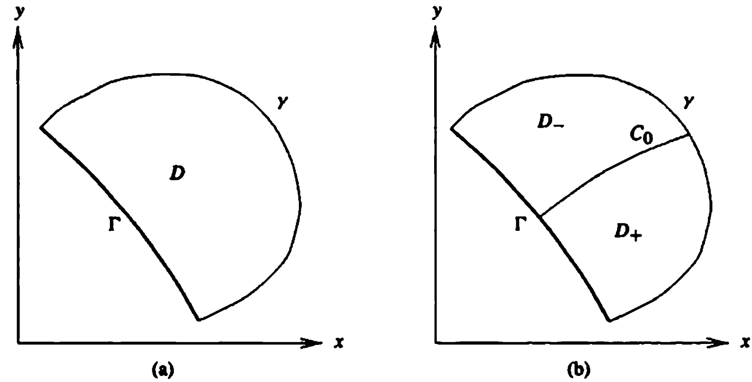


Fig. 1.8 Regions of integration for weak solutions: (a) continuous solutions; (b) discontinuous solutions.

$$\int_\Gamma u\psi(a\,\mathrm{d}y - b\,\mathrm{d}x) \equiv \iint_D u\left(\frac{\partial}{\partial x}(a\psi) + \frac{\partial}{\partial y}(b\psi) + \alpha\psi\right)\mathrm{d}x\,\mathrm{d}y. \tag{1.22}$$

Now (1.22) is an identity which can make sense even when u is not differentiable. Hence we define a *weak solution* of (1.20), with boundary data given on Γ, as a function u satisfying (1.22) for all test functions[8] ψ. The key point here is that this definition does not require u to possess derivatives at all points, only that the integrals exist; the onus of differentiability has been transferred from u to ψ. If, however, u is differentiable everywhere in a subdomain D_0, then, by choosing suitable test functions and reversing the steps in the argument above, it is easily shown that u satisfies (1.20) everywhere in D_0. A similar remark applies if u has discontinuous derivatives across a characteristic. However, if u itself is discontinuous across an open curve C_0, as in Fig. 1.8(b), then the application of Green's theorem to regions D^+ and D^- separated by C_0 gives, after subtraction,

$$\int_{C_0} \psi\,[u]_-^+\,(a\,\mathrm{d}y - b\,\mathrm{d}x) = 0. \tag{1.23}$$

Since $[u]_-^+ \neq 0$ and ψ is arbitrary, $\mathrm{d}y/\mathrm{d}x = b/a$ and C_0 is a characteristic projection.

For quasilinear problems this procedure can only be used if the equation is such that its principal part, i.e. the terms involving the highest derivatives, is in *divergence form* $\partial P/\partial x + \partial Q/\partial y$, so that Green's theorem can be applied. In this case,

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} = c, \tag{1.24}$$

[8] In some cases, such as models for gas dynamics, this definition can be motivated physically; see [30].

where P, Q and c are differentiable functions of x, y and u, and the equation is called an *inhomogeneous conservation law* (or, when c = 0, simply a *conservation law*).

The appropriate weak solution is then defined to be a function u such that

$$\iint_D \left(P\frac{\partial\psi}{\partial x} + Q\frac{\partial\psi}{\partial y} + c\psi\right) dx\, dy = \int_\Gamma \psi\,(P\, dy - Q\, dx) \tag{1.25}$$

for all admissible test functions ψ, and it is again easily verified that if u is differentiable it must satisfy (1.24). As in (1.23), if u is discontinuous across C_0,

$$\int_{C_0} \psi\left([P]_-^+\, dy - [Q]_-^+\, dx\right) = 0, \tag{1.26}$$

and, as ψ is arbitrary, this implies that the slope of the shock at C_0 is given by the *Rankine–Hugoniot condition*[9]

$$\frac{dy}{dx} = \frac{[Q]_-^+}{[P]_-^+}. \tag{1.27}$$

This is no longer a characteristic projection in general, but as $[u]_-^+ \to 0$ it tends to the characteristic projection defined by $dy/dx = (\partial Q/\partial u)/(\partial P/\partial u)$.[10]

Example 1.4 Solve the following ((1.14) again):

$$\frac{\partial u}{\partial x} + u\frac{\partial u}{\partial y} = 0, \quad u(0,y) = f(y).$$

By a simple generalisation of the discussion following (1.15), this problem fails to have a classical solution for all $x > 0$ whenever there are values of y for which $f'(y) < 0$. It does, however, have a weak solution defined by (1.25) with $P = u$, $Q = \frac{1}{2}u^2$ and $c = 0$. Then the possible curves C_0 separating domains in which classical solutions exist have slope such that

$$\frac{dy}{dx} = \frac{\left[\frac{1}{2}u^2\right]_-^+}{[u]_-^+} = \frac{u^+ + u^-}{2}.$$

Now comes the catch alluded to on p. 22: different weak solutions may arise from the same classical problem. Thus, in this example, if the partial differential equation were multiplied by u, so that $P = \frac{1}{2}u^2$ and $Q = \frac{1}{3}u^3$, the definition of the weak solution would be different and would lead to curves C_0 with slopes such that

$$\frac{dy}{dx} = \frac{\left[\frac{1}{3}u^3\right]_-^+}{\left[\frac{1}{2}u^2\right]_-^+} \neq \frac{u^+ + u^-}{2}.$$

[9]This condition can also be derived from a simple 'box' argument in which the jump in density across a small box centred on the shock is balanced by the jump in flux, much as in the derivation of mass conservation in §1.1. However, this argument is less easy to apply when there are more dependent variables.

[10]Note that the curve on which the solution of (1.11) is singular is not a characteristic either, but it has a slope which tends to that of a characteristic as $x \to \infty$.

Thus, in generalising the classical definition of a solution to the Cauchy problem, which leads to non-existence when shocks occur, we have lost uniqueness; additional conditions are needed to retrieve this property. We can see this more explicitly by returning to (1.14) with the initial data

$$u(0,y) = \begin{cases} 0, & y < 0, \\ 1, & y > 0. \end{cases}$$

A solution for which u is continuous but its derivatives are not is

$$u = \begin{cases} 0, & y < 0, \\ y/x, & 0 \leqslant y \leqslant x, \\ 1, & x < y; \end{cases}$$

the discontinuities propagate on the characteristic projections $y = 0$ and $y = x$. Thus it is a weak solution. However, another weak solution is

$$u = \begin{cases} 0, & y < x/2, \\ 1, & y > x/2, \end{cases}$$

since it satisfies the shock condition (1.27) on $y = x/2$. Thus there is more than one weak solution to this Cauchy problem, and in fact many such discontinuous weak solutions can be constructed. The fears expressed on p. 22 have been realised and some constraint or 'filter' is needed to restore uniqueness. There are several possibilities, three of which will be discussed at the end of Chapter 2.

*1.8 More independent variables

We conclude with some brief comments about the extension of the ideas of this chapter to scalar first-order quasilinear equations with more than two independent variables, which arise in several applications. For example, when a thin layer of a viscous liquid flows over a given curved surface $z = F(x,y)$, where z is vertically upwards, the appropriate generalisation of (1.1) is

$$\frac{\partial h}{\partial t} - \frac{1}{3}cA\left(\frac{\partial}{\partial x}\left(Ah^3\frac{\partial F}{\partial x}\right) + \frac{\partial}{\partial y}\left(Ah^3\frac{\partial F}{\partial y}\right)\right) = 0; \tag{1.28}$$

here $A = \left(1 + (\partial F/\partial x)^2 + (\partial F/\partial y)^2\right)^{-1/2}$, c is as in (1.1), and the thickness of the fluid layer measured normal to the substrate, h, is now a function of x, y and t.

The general equation has the form

$$\sum_{i=1}^{m} a_i \frac{\partial u}{\partial x_i} = c \quad \text{for } m > 2, \tag{1.29}$$

where a_i and c are functions of x_i and u. The appropriate generalisation of the Cauchy data (1.3) is to prescribe u on a surface of dimension $m-1$ and, to do this, a change of style is appropriate compared to that used in §1.3. With m coordinates to deal with, it is worth paying the price of representing the surface as a level set of a function, say $\phi(x_1, \ldots, x_m) = 0$, rather than selecting one preferred variable x_m as a function of the others.

We know u on the surface $\phi = 0$, which determines the $m-1$ tangential derivatives; together with the partial differential equation, this is in general sufficient to determine all the first derivatives $\partial u/\partial x_i$ on $\phi = 0$. To see whether this works analytically, we change from (x_i) to new variables (ϕ, ψ_i), $i = 1, \ldots, m-1$, where ψ_i are functions which are independent of ϕ. Then, by the chain rule,

$$\left(\sum_{i=1}^{m} a_i \frac{\partial \phi}{\partial x_i}\right) \frac{\partial u}{\partial \phi} + \mathbf{a} \cdot \nabla u = c, \tag{1.30}$$

where

$$\mathbf{a} \cdot \nabla = \sum_{j=1}^{m-1} \sum_{i=1}^{m} \frac{\partial \psi_j}{\partial x_i} \left(a_i \frac{\partial}{\partial \psi_j}\right).$$

Hence $\partial u/\partial \phi$, the derivative of u normal to $\phi = 0$, can be found as long as

$$\sum_{i=1}^{m} a_i \frac{\partial \phi}{\partial x_i} \neq 0. \tag{1.31}$$

Following (1.5), we say that $\phi = 0$, which is a surface of dimension $m-1$, is a *characteristic surface* of (1.29) if

$$\sum_{i=1}^{m} a_i \frac{\partial \phi}{\partial x_i} = 0. \tag{1.32}$$

We now have what seems to be a partial differential equation for ϕ, although we will not need to solve it as such.[11] In fact, since (1.32) only holds on $\phi = 0$, it is effectively an $(m-1)$-variable partial differential equation, albeit a nonlinear one (see Exercise 1.16).

It is easy to see that this approach to characteristics coincides with the definition that a surface is only characteristic if it can sustain a jump in $\partial u/\partial x_i$ across it. Indeed, subtracting (1.30) on either side of $\phi = 0$ gives

$$\left(\sum_{i=1}^{m} a_i \frac{\partial \phi}{\partial x_i}\right) \left[\frac{\partial u}{\partial \phi}\right]_{\phi=0-}^{\phi=0+} = 0.$$

However, the way in which such discontinuities propagate in $\phi = 0$ is a non-trivial generalisation of the corresponding result when $m = 2$, namely that a discontinuity at one point in the Cauchy data simply propagates along the characteristic projection through that point. First, assume for simplicity that the a_i are constant

[11] In Chapter 8 we will see that the solution of equations such as (1.32) can always be reduced to the solution of some ordinary differential equations.

and $c = 0$. This means that the characteristics are hyperplanes and so we can take ϕ to be a linear combination of the x_i. Then, differentiating (1.30) with respect to ϕ and subtracting on either side of $\phi = 0$, we obtain

$$\left(\sum_{i=1}^{m} a_i \frac{\partial \phi}{\partial x_i}\right) \left[\frac{\partial^2 u}{\partial \phi^2}\right]_{\phi=0-}^{\phi=0+} + (\mathbf{a} \cdot \nabla) \left[\frac{\partial u}{\partial \phi}\right]_{\phi=0-}^{\phi=0+} = 0. \tag{1.33}$$

Hence, on a characteristic $\phi = 0$,

$$(\mathbf{a} \cdot \nabla) \left[\frac{\partial u}{\partial \phi}\right]_{\phi=0-}^{\phi=0+} = 0.$$

Even when a_i and c depend on x_i and u, it is easy to see that the right-hand side of (1.33) is replaced by a term that is simply proportional to $[\partial u/\partial \phi]_{\phi=0-}^{\phi=0+}$, as in (1.18). Thus we have the interesting result that the coefficients a_i define a direction in the characteristic surface, called the ***bicharacteristic direction***, in which discontinuities propagate. Any jump in the Cauchy data at one point of the initial surface only propagates along the one-dimensional bicharacteristic through that point. Moreover, it is easy to see that the $(m-1)$-dimensional characteristic surfaces are each made up of an $(m-2)$-parameter family of bicharacteristic curves (see Fig. 1.9 for the case $m = 3$). Of course, bicharacteristics are simply characteristics when $m = 2$; for $m > 2$, they can be used to construct the solution parametrically by solving the ***bicharacteristic equations***

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = a_i \quad \text{with} \quad \frac{\mathrm{d}u}{\mathrm{d}t} = c, \tag{1.34}$$

together with the Cauchy data

$$u = u_0(\mathbf{s}), \quad x_i = x_{i0}(\mathbf{s}), \quad \mathbf{s} = (s_1, \ldots, s_{m-1}).$$

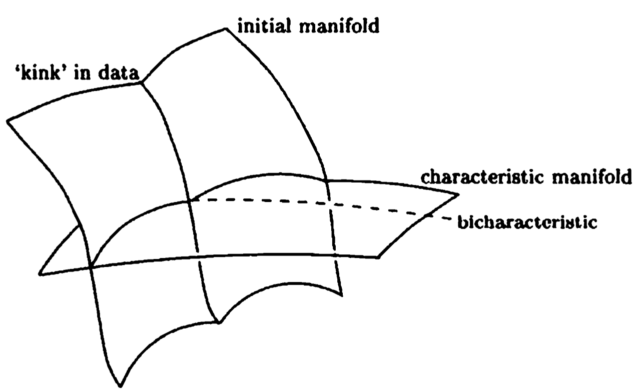


Fig. 1.9 Characteristic surface and bicharacteristic curve for (1.29).

Example 1.5 Solve

$$x\frac{\partial u}{\partial x}+y\frac{\partial u}{\partial y}+z\frac{\partial u}{\partial z}=1$$

with

$$u=0 \quad \text{on } x+y+z=1,$$

i.e.

$$u_0(s_1,s_2)=0, \quad x_0(s_1,s_2)=s_1, \quad y_0(s_1,s_2)=s_2, \quad z_0(s_1,s_2)=1-s_1-s_2.$$

The bicharacteristic equations are

$$\dot{x}=x, \quad \dot{y}=y, \quad \dot{z}=z, \quad \dot{u}=1,$$

and the parametric solution is then

$$x=s_1\mathrm{e}^t, \quad y=s_2\mathrm{e}^t, \quad z=(1-s_1-s_2)\mathrm{e}^t, \quad u=t.$$

Non-parametrically, the solution is $u=\log(x+y+z)$ and the characteristic surfaces are given by $x\,\partial\phi/\partial x+y\,\partial\phi/\partial y+z\,\partial\phi/\partial z=0$, which, by a generalisation of the method described on p. 16, is easily seen to have the general solution $\phi=F(x/y,x/z)$. Thus, the characteristic surfaces are cones of arbitrary cross-section through the origin, and the bicharacteristics are the generators of these cones.

1.9 Postscript

It is helpful to have a catalogue of the number of arbitrary constants and functions that we might expect to appear in the general solutions of differential equations, and we conclude this chapter with such a list. Functions denoted by F are determined by the equation and those denoted by G are arbitrary.

For *ordinary differential equations* for functions $u(x)$, we have the following.

First-order: The general solution is

$$F(u,x,G_1)=0,$$

where G_1 is an arbitrary constant.

Second-order: The general solution is

$$F(u,x,G_1,G_2)=0,$$

where $G_{1,2}$ are constant. The pattern continues in this way for higher orders.

For *partial differential equations*, intuition suggests that the solutions depend on arbitrary functions as follows.

First-order, two independent variables: The general solution is

$$F_0(u,x,y,G_1(F_1))=0,$$

where G_1 is an arbitrary function and $F_1=F_1(u,x,y)$.

First-order, three independent variables: The general solution is

$$F_0(u, x, y, z, G_1(F_1, F_2)) = 0,$$

where G_1 is arbitrary and $F_{1,2} = F_{1,2}(u, x, y, z)$.

Second-order, two independent variables: The general solution is

$$F_0\,(u, x, y, G_1(F_1), G_2(F_2)) = 0,$$

where $G_{1,2}$ are arbitrary and $F_{1,2} = F_{1,2}(u, x, y)$.

Second-order, three independent variables: The general solution is

$$F_0(u, x, y, z, G_1(F_1, F_2), G_2(F_3, F_4)) = 0,$$

where $G_{1,2}$ are arbitrary and $F_{1,2,3,4} = F_{1,2,3,4}(u, x, y, z)$.

Again, the pattern continues for higher orders and more independent variables. However, this list can be misleading should any of the functions or variables be complex, as we shall see in Chapters 3 and 5.

Exercises

1.1. Consider the proof-reading model of §1.1 with N errors initially. Show that the solution is

$$p(x,t) = \left(1 + (x-1)\mathrm{e}^{-\mu t}\right)^N.$$

Does the result agree with your experience of reading this book, which has been proof-read 33 times over two editions?

1.2. Suppose that $\rho(x,t)$ is the number density of cars per unit length along a road, x being distance along the road, and let $u(x,t)$ be their velocity (overtaking is illegal). Assuming that no cars enter or leave the road, show that, if $a(t)$ and $b(t)$ are the positions of any two cars (so that $\mathrm{d}a/\mathrm{d}t = u(a,t)$ and $\mathrm{d}b/\mathrm{d}t = u(b,t)$), then

$$\int_{a(t)}^{b(t)} \rho(x,t)\,\mathrm{d}x$$

is independent of time. Deduce that, if ρ and u are sufficiently well behaved,

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0.$$

Suppose further that u is a known decreasing function of ρ (why is this realistic?). Show that information propagates through the traffic at a velocity $\mathrm{d}(\rho u)/\mathrm{d}\rho < u$.

1.3. (i) Suppose that you can spot one integral $f(x,y,u) = k =$ constant of the characteristic equations

$$\frac{\dot{x}}{a(x,y,u)} = \frac{\dot{y}}{b(x,y,u)} = \frac{\dot{u}}{c(x,y,u)}$$

and that you can solve for $u = F(x,y,k)$. Assuming you are also lucky enough to be able to integrate $\mathrm{d}y/\mathrm{d}x = b(x,y,F)/a(x,y,F)$ with $k =$ constant, explain how to find a second integral of the characteristic equations.

(ii) The function a is said to be *homogeneous* if, for all λ,

$$a(\lambda x, \lambda y, \lambda u) \equiv \lambda^{\alpha} a(x,y,u)$$

for some number α. Show that, if a, b and c are all homogeneous functions with the same value of α, the two characteristic equations can be reduced to a single first-order scalar differential equation.

Hint. Note that a, b and c are functions only of, say, u/x and y/x, multiplied by x^{α}; write $u = xv(\xi)$, where $\xi = y/x$, to obtain an equation for $\mathrm{d}v/\mathrm{d}\xi$ just as a function of v and ξ.

1.4. Consider equation (1.12), for which two first integrals of the characteristic equations are given by $f(x,y,u) = x + y + u =$ constant and $g(x,y,u) = xyu =$ constant, and the general solution is $G(f,g) = 0$. Suppose that the initial data is $x_0(s) = y_0(s) = u_0(s) = s$. Show that one choice for G is $G(f,g) = (f/3)^3 - g$, and hence that the solution is given implicitly by $((x+y+u)/3)^3 - xyu = 0$.

1.5. Integrate the characteristic equations to show that the solution of

$$y\frac{\partial u}{\partial x} - 2xy\frac{\partial u}{\partial y} = 2xu,$$

with $u = y^3$ when $x = 0$ and $1 \leqslant y \leqslant 2$, is

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

What is the domain of definition of the solution in $y > 0$?

1.6. Integrate the characteristic equations to show that the solution of

$$x^3\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y},$$

with

$$u = \frac{1}{1+x^2} \quad \text{on } y = 0,\ -\infty < x < \infty,$$

is

$$u = \frac{1 - 2x^2 y}{1 + x^2 - 2x^2 y}.$$

Show that the solution is not defined in $y > 1/2x^2$ despite the fact that the data is prescribed for all x.

1.7. Suppose that

$$xu\frac{\partial u}{\partial x} - yu\frac{\partial u}{\partial y} = x^2 - y^2$$

and that $u = f(x)$ on $x = y$. Show that when

(i)	$f(x) = 0,$	$u = \pm(x - y),$
(ii)	$f(x) = \sqrt{2}x,$	$u = \pm\sqrt{x^2 + y^2},$
(iii)	$f(x) = x,$	$u = \pm\sqrt{x^2 - xy + y^2},$

for $xy > 0$, and determine which sign should be taken in the first quadrant and which in the third. Explain why the solutions are non-unique in case (i) and why the solution cannot be defined when $xy < 0$. Describe the surfaces $z = u(x, y)$ geometrically. What change of variable would make these problems easier?

1.8. Derive the parametric solution of

$$(x + u)\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = u + y^2$$

in the form

$$y = y_0(s)e^t, \quad u = \left(u_0(s) - y_0^2(s)\right)e^t + y_0^2(s)e^{2t},$$
$$x = \left(x_0(s) - y_0^2(s)\right)e^t + \left(u_0(s) - y_0^2(s)\right)te^t + y_0^2(s)e^{2t}.$$

Suppose that $u = x$ on $y = 1, -\infty < x < \infty$. Show that

$$u(x, y) = \frac{x - y^2}{1 + \log y} + y^2.$$

What is the domain of definition of u?

1.9. Show that if

$$-y\frac{\partial u}{\partial x} + x\frac{\partial u}{\partial y} = f(x, y)$$

then, away from $x = y = 0$,

$$u = \int_0^\theta f(r\cos\theta, r\sin\theta)\,d\theta + F(r), \tag{1.35}$$

where $x = r\cos\theta$, $y = r\sin\theta$ and $F(r)$ is arbitrary. Now suppose that $\int_0^{2\pi} f(r\cos\theta, r\sin\theta)\,d\theta = 0$. Show that u can be defined for all $(x, y) \neq (0, 0)$. Suppose further that $f(x, y)$ is analytic at $(0, 0)$, so that it has a Taylor series expansion which converges to f in a neighbourhood of the origin. Show that, if u is also analytic at the origin, then

$$u = \frac{1}{2}\left(\int_\alpha^\theta + \int_{\pi+\alpha}^\theta f(r\cos\theta, r\sin\theta)\,d\theta\right) + G(r^2),$$

where $G(z)$ is analytic at $z = 0$ and α is arbitrary.

Hint. Try $f = r\sin\theta$ in (1.35).

(We are grateful to S. Dobrokhotov for this example.)

* 1.10. Take $u = 1-\rho$ in Exercise 1.2 for $0 \leqslant \rho \leqslant 1$. How might you interpret $\rho = 1$? Show that u and ρ are constant on the characteristics

$$\frac{dx}{dt} = 1 - 2\rho,$$

and show that the Rankine–Hugoniot condition for the speed of a shock $x = S(t)$ is

$$\frac{dS}{dt} = \frac{[\rho(1-\rho)]_-^+}{[\rho]_-^+}.$$

A queue is building up at a traffic light $x = 1$ so that, when the light turns to green at $t = 0$,

$$\rho(x,0) = \begin{cases} 0, & x < 0 \text{ and } x > 1, \\ x, & 0 < x < 1. \end{cases}$$

Show that initially the characteristics, which are straight, are

$$\begin{array}{ll} x - s = t & \text{in } x < t \text{ and } x > t+1, \text{ where } \rho = 0, \\ x - s = (1-2s)t & \text{in } t < x < 1-t, \text{ where } \rho = s, \\ x - 1 = (1-2s)t & \text{in } 1-t < x < 1+t, \text{ where } \rho = s = (t-x+1)/(2t). \end{array}$$

Deduce that a collision first occurs at $x = 1/2$ when $t = 1/2$, and that thereafter there is a shock such that

$$\frac{dS}{dt} = \frac{S+t-1}{2t}.$$

1.11. Show that the solution of

$$\frac{\partial u}{\partial x} + u\frac{\partial u}{\partial y} = 1,$$

with $u = s/2$ on $x = y = s$ for $0 \leqslant s \leqslant 1$, is

$$u = \frac{4x - 2y - x^2}{2(2-x)},$$

and that the characteristics are

$$y = \frac{x^2}{2} + c(2-x), \quad c = \text{constant}.$$

What is the domain of definition of the solution?

$*$ 1.12. Show that, if

$$\frac{\partial u}{\partial x} + u\frac{\partial u}{\partial y} = 0$$

for $x > 0$ with

$$u(0, y) = \begin{cases} 0, & y \leqslant 0 \text{ or } y \geqslant 1, \\ y(1-y), & 0 \leqslant y \leqslant 1, \end{cases}$$

then, for $0 \leqslant s \leqslant 1$, the characteristics are

$$y - s = s(1-s)x$$

with

$$u = s(1-s).$$

Show also that

$$u^2x^2 + u(1 + x - 2xy) + y^2 - y = 0,$$

and show that u is continuous for small enough x. The envelope of the characteristics is found by differentiating $y - s = s(1-s)x$ partially with respect to s and eliminating s; show that this gives $4xy = (x+1)^2$ and deduce that a shock forms at $x = y = 1$.

$*$ 1.13. Paint flowing down a wall has thickness $u(x, t)$ satisfying

$$\frac{\partial u}{\partial t} + u^2\frac{\partial u}{\partial x} = 0 \quad \text{for } t > 0.$$

Show that the characteristics are straight and that the Rankine–Hugoniot condition on a shock $x = S(t)$ is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{[\frac{1}{3}u^3]_-^+}{[u]_-^+}.$$

A stripe of paint is applied at $t = 0$ so that

$$u(x, 0) = \begin{cases} 0, & x < 0 \text{ or } x > 1, \\ 1, & 0 < x < 1. \end{cases}$$

Show that, for small enough t,

$$u = \begin{cases} 0, & x < 0, \\ (x/t)^{1/2}, & 0 < x < t, \\ 1, & t < x < S(t), \\ 0, & S(t) < x, \end{cases}$$

where the shock is $x = S(t) = 1 + t/3$. Explain why this solution changes at $t = 3/2$, and show that thereafter

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{S}{3t}.$$

***1.14.** Integrate along the bicharacteristics to show that, if

$$u\frac{\partial u}{\partial x}+z\frac{\partial u}{\partial y}+\frac{\partial u}{\partial z}=y,$$

with $u=x$ on $z=0$, $-\infty<x,y<\infty$, then

$$z=t,\quad y=\frac{t^2}{2}+s_2,\quad u=\frac{t^3}{6}+s_2t+s_1,\quad x=\frac{t^4}{24}+\frac{s_2t^2}{2}+s_1(t+1).$$

Deduce that

$$u=\frac{24x-12yz^2+5z^4}{24\,(1+z)}+yz-\frac{1}{3}z^3$$

and that the domain of definition is bounded by $z=-1$.

***1.15.** Show that the projections of the bicharacteristics of (1.28) onto the (x,y) plane are lines of steepest descent of the surface $z=F(x,y)$, i.e. that they are orthogonal to the level curves of F.

***1.16.** Suppose that (1.32) is satisfied on $\phi=0$ and that the equation $\phi=0$ can be solved locally for $x_m=\psi(x_1,\dots,x_{m-1})$. Show that the differential equation

$$\sum_{i=1}^{m-1}a_i\frac{\partial\psi}{\partial x_i}-a_m=0,$$

where $a_i=a_i(x_1,\dots,x_{m-1},\psi)$, holds in the space of $(x_1,\dots,x_{m-1})$.

***1.17.** Suppose that $u(x,y)$ is such that $\partial u/\partial x=0$ with

$$u(0,y)=\begin{cases}-1, & y<0,\\ +1, & y>0.\end{cases}$$

Now let the partial differential equation for u be replaced by

$$\frac{\partial u}{\partial x}=\epsilon\frac{\partial^2 u}{\partial y^2},$$

for small positive ϵ. Verify that a solution of this equation is

$$u(x,y)=\frac{2}{\sqrt{\pi}}\int_0^{y/2\sqrt{\epsilon x}}\mathrm{e}^{-s^2}\,\mathrm{d}s$$

(this will be derived in Chapter 6). Show that, as $\epsilon\to 0$,

$$u\to\begin{cases}-1, & y<0,\\ 1, & y>0,\end{cases}$$

for $x\geqslant 0$, and that this result is the same as that obtained by requiring u to be discontinuous only on a characteristic. (This way of smoothing a shock will be studied further in Chapter 2.)

2

First-order quasilinear systems

2.1 Motivation and models

When we use vector systems of partial differential equations, we can model many more physical situations than when we are restricted to the scalar case. For example, as well as looking at the simple kinematic wave models of Chapter 1, we can study situations which allow simultaneous wave propagation both backwards and forwards. A famous example is that of *shallow water waves* where gravity and fluid inertia balance each other in a shallow layer of water with a free surface, of depth $h(x,t)$, flowing above a horizontal bed along the x axis (Fig. 2.1(a)). Let the fluid density be a constant ρ. Then if we assume the pressure p is nearly hydrostatic, and set it equal to $\rho g(h-y)$, the horizontal force balance[12] for a river, whose bed is the x axis, flowing nearly unidirectionally with velocity $u(x,t)$ is

$$\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)=-\frac{\partial p}{\partial x}=-\rho g\frac{\partial h}{\partial x}. \tag{2.1}$$

Also, as in the paint model of §1.1, conservation of mass gives

$$\frac{\partial h}{\partial t}+\frac{\partial}{\partial x}(hu)=0. \tag{2.2}$$

We thus have a two-by-two system for u and h.

The shallow water equations can be shown to be (see Exercise 2.1) a special case of the following three-by-three system of one-dimensional unsteady gas dynamics (Fig. 2.1(b)):

$$\frac{\partial \rho}{\partial t}+\frac{\partial}{\partial x}(\rho u)=0,\quad \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}, \tag{2.3}$$

$$\rho\left(\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}\right)\left(\frac{p}{(\gamma-1)\rho}+\frac{1}{2}u^2\right)+\frac{\partial}{\partial x}(pu)=0, \tag{2.4}$$

for the pressure p, the density ρ (which now varies) and the velocity u, where γ is a constant greater than unity; these equations represent mass, force and energy balances, respectively. These are also the balances needed to model two-dimensional steady gas dynamics with velocity (u,v) as the four-by-four system

[12] In a fluid flow the acceleration is not just the time derivative of the velocity; the convective term $u\,\partial u/\partial x$ needs to be added to account for the fact that the fluid accelerates as it moves from place to place.

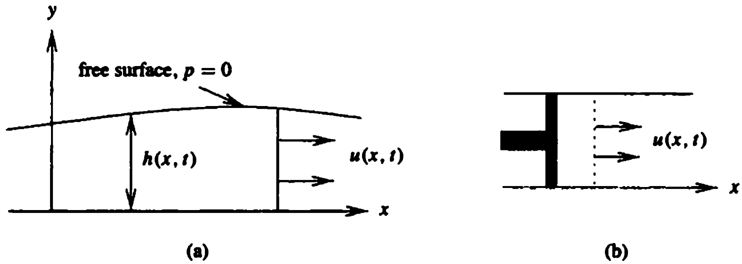


Fig. 2.1 (a) Open channel shallow water flow; (b) one-dimensional unsteady gas flow driven by a piston.

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) = 0, \tag{2.5}$$

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \quad u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial y}, \tag{2.6}$$

$$\rho\left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\left(\frac{p}{(\gamma-1)\rho} + \frac{1}{2}(u^2+v^2)\right) + \frac{\partial}{\partial x}(pu) + \frac{\partial}{\partial y}(pv) = 0. \tag{2.7}$$

In both gasdynamics models the final equation expresses the fact that all the work done on the gas by the pressure forces goes into changing the heat content and kinetic energy.[13] A further generalisation is to allow for chemical reactions to occur in the gas, in which case (2.4) and (2.7) have non-zero right-hand sides involving, in general, the reactant concentrations for which suitable rate equations must be written down. Later in the chapter, we will consider a simple case when these right-hand sides are localised in space, which will enable us to model the dramatic 'detonations' that can sometimes accompany shock waves.

As they stand, (2.5)–(2.7) are a complicated quasilinear system, so it is instructive to consider the simpler model obtained by 'linearising' about the constant solution $u = U$, $v = 0$, $p = p_0$, $\rho = \rho_0$. We will discuss such linearisations in more detail in Chapter 7, but it can be shown that (2.7) often implies p/ρ^γ = constant $= p_0/\rho_0^\gamma$ and thus that, if we write $u = U + \bar{u}$, $v = \bar{v}$, $p = p_0 + \bar{p}$, $\rho = \rho_0 + \bar{\rho}$ and neglect squares of barred quantities, we obtain

$$\rho_0\left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y}\right) + U\frac{\partial \bar{\rho}}{\partial x} = 0, \tag{2.8}$$

$$U\frac{\partial \bar{u}}{\partial x} + \frac{a_0^2}{\rho_0}\frac{\partial \bar{\rho}}{\partial x} = 0, \tag{2.9}$$

$$U\frac{\partial \bar{v}}{\partial x} + \frac{a_0^2}{\rho_0}\frac{\partial \bar{\rho}}{\partial y} = 0, \tag{2.10}$$

where $a_0^2 = \gamma p_0/\rho_0$ (see Exercise 2.2); a_0 will later be seen to be the speed of sound in the gas.

[13]We have also assumed that the gas obeys the so-called *ideal gas law* in which its temperature is proportional to p/ρ.

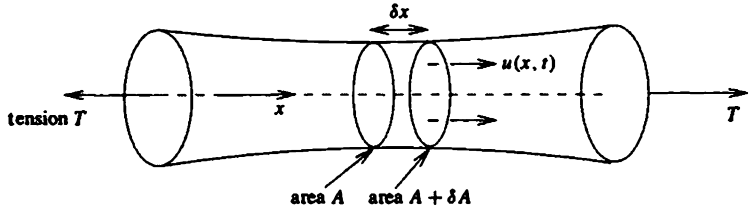


Fig. 2.2 Fibre-drawing.

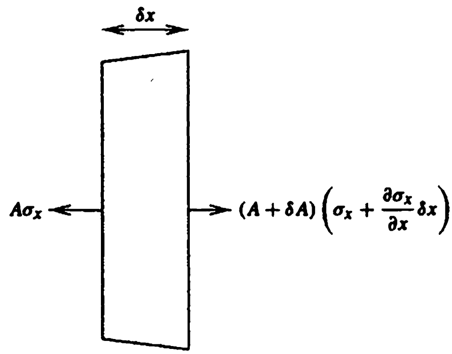


Fig. 2.3 Fibre-drawing: force balance.

Less well-studied systems of first-order equations arise in many industrial processes. A fertile source is in the glass industry where, for example, the nearly unidirectional drawing of an optical fibre of cross-sectional area $A(x,t)$ with velocity $u(x,t)$ can be modelled by what is called an 'extensional flow' (Fig. 2.2). This is one in which the principal viscous resistance is not one of shearing (as in the paint model of Chapter 1) but of a *normal* stress σ_x proportional to $\partial u/\partial x$ giving a force proportional to $A\,\partial u/\partial x$ (Fig. 2.3). When we neglect the inertia of the glass and integrate the momentum balance $\partial\sigma_x/\partial x = 0$, we obtain

$$cA\frac{\partial u}{\partial x} = T(t), \tag{2.11}$$

where T is the tension in the fibre; the constant c is proportional to the *viscosity* of the glass, which measures its resistance to shearing and extension.[14] Conservation of mass implies, as usual,

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial x}(Au) = 0, \tag{2.12}$$

and so we have a slightly unconventional first-order system in which T is to be determined by the boundary conditions.

[14]In the paint model of §1.1, the counterpart of this constant is inversely proportional to the viscosity.

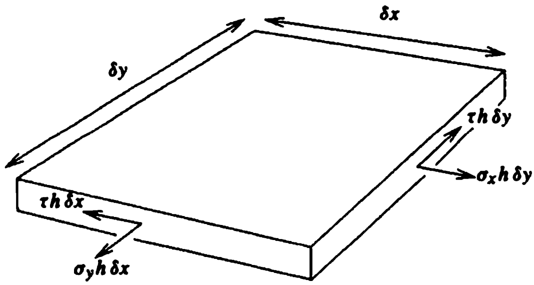


Fig. 2.4 Glass sheet stretching: force balance.

An even more interesting configuration occurs when we consider the extension of a sheet of glass of thickness $2h(x, y, t)$, symmetric about the (x, y) plane (Fig. 2.4). Then two-dimensional shearing of the velocity (u, v) occurs in the (x, y) plane. When we integrate across the thickness of the sheet, the momentum balances for the normal and shear stresses σ_x and σ_y, and τ, respectively, are

$$\frac{\partial(h\sigma_x)}{\partial x} + \frac{\partial(h\tau)}{\partial y} = \frac{\partial(h\tau)}{\partial x} + \frac{\partial(h\sigma_y)}{\partial y} = 0. \tag{2.13}$$

The generalisations of the models of Fig. 1.1(b) and Fig. 2.3 can be shown to lead to forces

$$\begin{gathered} h\sigma_x = 2h\left(2\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \quad h\sigma_y = 2h\left(\frac{\partial u}{\partial x} + 2\frac{\partial v}{\partial y}\right), \\ h\tau = h\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \end{gathered} \tag{2.14}$$

where the constant viscosity has been set equal to unity for convenience. The mass conservation equation in (2.3) generalises to

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0, \tag{2.15}$$

leaving us with the six-dimensional system (2.13)–(2.15).

Other examples in process engineering include the following.

Flow of granular materials

A force balance in an inertia-free two-dimensional granular material in the absence of gravity is also simply (2.13) with $h = 1$ (Fig. 2.5). On the assumption that flow is taking place because the stresses are everywhere strong enough to overcome friction, we can find the simple 'yield criterion', which the stress must satisfy, from an analysis using Coulomb's law of friction. This just says that we need to ensure that at each point of a flowing granular material there is a 'slip plane', on an element of which the ratio of the shear (frictional) force to the normal force is equal to the coefficient of friction, say $\tan\phi$, while on all other planes the ratio is less than $\tan\phi$.

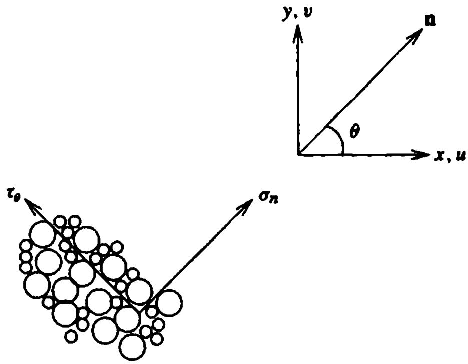


Fig. 2.5 Flow of a granular medium; σ_n is the normal stress and τ_θ the shear stress on a surface with normal **n**.

Although this condition is easy to state in words, the calculation of Exercise 2.3 is needed to translate it into the statement that

$$\tau = -\left(\frac{\sigma_x + \sigma_y}{2}\right) \sin\phi \sin\left(\tan^{-1}\left(\frac{2\tau}{|\sigma_x - \sigma_y|}\right)\right) \tag{2.16}$$

which, with (2.13), gives a closed model for σ_x, σ_y and τ.

Flow in packed beds and fluidised beds

There are many practical applications of systems of first-order partial differential equations in chemical engineering. Quite apart from models for reactors, about which we will say more in Chapter 6, almost any process involving heat or mass transfer from one material phase to another results in a model involving first-order derivatives in time and space.

The simplest situation occurs when the position of one material phase is known in advance, as, for example, in a packed bed of solid particles through which a fluid is being passed. Typically, we might be considering a regenerative heat exchanger in which a cold gas at temperature θ_g needs to be heated; this is done by passing the gas with speed U through stationary solid particulate whose temperature is θ_s, which itself may have been heated by a different hot gas. Thus, in a one-dimensional configuration, ignoring heat conduction, the only balance that needs to be written down is that between heat convection and heat transfer in the thin gas layer adjacent to each particle of the bed. For the gas, this yields an averaged heat balance in the form

$$\rho_g c_g \left(\frac{\partial \theta_g}{\partial t} + U\frac{\partial \theta_g}{\partial x}\right) = h(\theta_s - \theta_g), \tag{2.17}$$

and, for the particulate,

$$\rho_s c_s \frac{\partial \theta_s}{\partial t} = h(\theta_g - \theta_s), \tag{2.18}$$

where ρ and c refer to density and specific heat, respectively, and h is an experimentally determined 'heat transfer coefficient'.

More complicated situations occur when both phases can move as, say, in a fluidised bed, in which the gravity force acting to immobilise the particulate phase can be overcome by the drag force of fluid flowing vertically from a porous 'distributor' at the bottom of the bed. It is notoriously difficult to model the mechanics of this process but, if we are only interested in modelling the mixing properties of the fluid in the bed, we can note from observation that in many situations the gas flows in two modes. In addition to intergranular porous medium flow as assumed in the regenerator, large, often kidney-shaped, bubbles of almost pure gas rise continually from the distributor to the surface of the bed. Denoting the gas concentrations in these modes by c_a and c_b, respectively, we find mass balance laws

$$f_a\left(\frac{\partial c_a}{\partial t} + v_a\frac{\partial c_a}{\partial x}\right) + k(c_a - c_b) = 0, \tag{2.19}$$

$$f_b\left(\frac{\partial c_b}{\partial t} + v_b\frac{\partial c_b}{\partial x}\right) + k(c_b - c_a) = 0, \tag{2.20}$$

where f_a and f_b are the volume fractions occupied by pore space and bubbles, respectively, v_a and v_b are the gas velocities in the two modes, and k is now a mass transfer coefficient; all these parameters are assumed constant. This is the so-called *Van Deemter model* for mixing in a bubbling fluidised bed, and it is clearly a generalisation of (2.17) and (2.18).

It is not surprising that so many models of partial differential equations comprise combinations of first-order partial derivatives. This is because all the laws of balance of mass, momentum and energy are typically expressed as

$$\frac{\partial}{\partial t}(\text{density}) + \nabla\cdot(\text{flux}) = \text{source},$$

and furthermore the constitutive laws describing the mechanical or thermodynamic behaviour of many materials only involve first derivatives of certain physical variables. On top of this, problems involving electromagnetism demand the incorporation of Maxwell's equations, which in their basic form are relationships between the first partial derivatives of the electric and magnetic fields in space and time. The only common physical situation where higher-order derivatives occur naturally at this modelling stage is quantum mechanics, about which we will say more in Chapter 8.

With this powerful motivation, we now describe a framework within which first-order systems with two independent variables may be considered. Alas, too few of the examples listed above (or indeed of any nonlinear system of partial differential equations) can be analysed as fully as those in Chapter 1, and those for which substantial progress can be made can often be most conveniently written as higher-order scalar equations. Hence this chapter does not offer any great insight into the precise behaviour of the solution of any particular partial differential equation system; rather it describes how all such equations should be assessed before any detailed analytical or numerical investigation is undertaken.

2.2 Cauchy data and characteristics

We first attempt to generalise the ideas of Chapter 1 to a real first-order system of dimension two or greater and, until §2.6, we only consider the case of two independent variables (x, y). We begin by considering two-by-two systems, for which we seek a vector function $\mathbf{u}$ with components u_1 and u_2 satisfying

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{c}, \tag{2.21}$$

where $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are 2×2 matrices and $\mathbf{c}$ is a vector with two components, all of whose entries and components are functions of x, y, u_1 and u_2. A three-dimensional geometrical interpretation is that we are looking for two surfaces $u_1 = u_1(x, y)$ and $u_2 = u_2(x, y)$; we expect appropriate boundary data to be that each surface passes through its own prescribed initial curve. Such a boundary condition can be written in parametric form as

$$\mathbf{u} = \mathbf{u}_0(s), \quad x = x_0(s), \quad y = y_0(s) \quad \text{for } s_1 \leqslant s \leqslant s_2.$$

This boundary data implies that the two partial derivatives of $\mathbf{u}$ must satisfy

$$\mathbf{u}_0' = x_0'\frac{\partial \mathbf{u}}{\partial x} + y_0'\frac{\partial \mathbf{u}}{\partial y}, \tag{2.22}$$

where $'$ again denotes $\mathrm{d}/\mathrm{d}s$. These partial derivatives are uniquely determined on the boundary curve by the four equations (2.21) and (2.22) provided that

$$\begin{vmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ x_0' & 0 & y_0' & 0 \\ 0 & x_0' & 0 & y_0' \end{vmatrix} \neq 0 \qquad \text{for } s_1 \leqslant s \leqslant s_2. \tag{2.23}$$

With $y_0' = \lambda x_0'$, we may subtract suitable multiples of the last two columns from the first two to obtain

$$\det(\mathbf{B} - \lambda\mathbf{A}) \neq 0. \tag{2.24}$$

The condition (2.24) clearly reduces to (1.5) in the one-dimensional case; it is also true for first-order systems when $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices. Now, in the scalar case, the methods of Chapter 1 were sufficient to reduce the problem to a well-behaved system of ordinary differential equations as long as (1.5) was satisfied. Hence, assuming Lipschitz continuity, the *scalar* problem is *well posed*, by which we mean that

- the solution exists;
- it is unique;
- it depends continuously on the data.

However, we will soon see that in the vector case $n > 1$, (2.24) is no longer sufficient to ensure well-posedness (the 'data' now being $\mathbf{A}$, $\mathbf{B}$, $\mathbf{c}$ and the Cauchy data for $\mathbf{u}$). Surprisingly, we will see that it does not even guarantee this property for linear problems, in which the kind of singularity development described in §§1.4 and 1.5 could not occur. Thus, great care is necessary in posing appropriate boundary conditions for the system (2.21).

In Chapter 1, a characteristic was defined by the characteristic ordinary differential equations (1.7), for which there is no obvious generalisation to higher dimensions. However, we can generalise the idea mentioned after (1.5), that if Cauchy data is given on a curve which is the projection of a characteristic then the partial derivatives are not uniquely defined on that curve. In a first-order scalar problem, the Cauchy data defines the partial derivative in the direction of the curve and it is the partial derivative normal to the curve that is ill-defined on a characteristic projection. We thus define a *characteristic* (projection) of a system by this property, and henceforth we will omit the word projection when discussing the characteristic projections of a system.[15]

With this definition, a characteristic of $\mathbf{A}\,\partial\mathbf{u}/\partial x + \mathbf{B}\,\partial\mathbf{u}/\partial y = \mathbf{c}$ is a curve $(x(t), y(t))$ such that this equation evaluated on the curve and

$$\dot{x}\frac{\partial\mathbf{u}}{\partial x} + \dot{y}\frac{\partial\mathbf{u}}{\partial y} = \dot{\mathbf{u}} \tag{2.25}$$

do not have unique solutions for $\partial\mathbf{u}/\partial x$ and $\partial\mathbf{u}/\partial y$, which means that the left-hand sides of these four equations are linearly dependent. This is just the calculation leading to (2.24) with s replaced by t; hence we say that $(x(t), y(t))$ is a characteristic if

$$\det(\mathbf{B}\dot{x} - \mathbf{A}\dot{y}) = 0. \tag{2.26}$$

This is a quadratic expression in $\mathrm{d}y/\mathrm{d}x = \dot{y}/\dot{x}$ and may therefore not give real characteristic directions at each point, unlike the situation in the scalar case. Hence, we already see a fundamental contrast with the cases considered in Chapter 1: the *change in type* (a concept that we will make precise in the next chapter) that occurs when the characteristics change from being real to complex is no less dramatic than the striking new phenomena that occur when an aircraft penetrates the sound barrier.[16] This can be seen by the following simple calculations on the equations (2.8)–(2.10), which is an example to which we will return repeatedly throughout the rest of the book. In this example, $\mathbf{u}$ has three components but the result (2.26) generalises trivially to any number of dimensions.

[15] An equivalent definition in the first-order scalar case was used in (1.17) to say that, if the first derivatives of u are smooth except for jump discontinuities, then this jump must be along a characteristic. We will return to this contrasting approach shortly; alas, the geometrical interpretation for a system of dimension two is in a space of dimension four, so it is less easy to visualise than it was for the scalar case.

[16] It also yet further handicaps any generalisation of the approach of Chapter 1, for complex characteristics would have to be considered in a four-dimensional real space, just for equations with two independent variables.

Example 2.1 (*Steady two-dimensional gas dynamics*) For the linearised gasdynamic model (2.8)–(2.10) we find

$$\mathbf{A} = \begin{pmatrix} \rho_0 & 0 & U \\ U & 0 & a_0^2/\rho_0 \\ 0 & U & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & \rho_0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a_0^2/\rho_0 \end{pmatrix}, \quad \text{where } \mathbf{u} = \begin{pmatrix} \bar{u} \\ \bar{v} \\ \bar{\rho} \end{pmatrix}.$$

Hence, $\det(\mathbf{B} - \lambda\mathbf{A}) = \lambda a_0^2 U(\lambda^2 + a_0^2 - U^2)$ and we have three real characteristics only if $U^2 > a_0^2$. When this inequality holds, the constant state about which we have linearised is *supersonic*, and the characteristics in such a flow are the streamlines of the unperturbed flow, $\mathrm{d}y/\mathrm{d}x = 0$, and the so-called 'Mach lines' $\mathrm{d}y/\mathrm{d}x = \pm(U^2 - a_0^2)^{1/2}$.

We will pursue the implications of the reality, or otherwise, of the characteristics more generally in Chapter 3. For the moment, we note that there is still more information buried in (2.25). Since (2.21) and (2.25) are linearly dependent along a characteristic, a further relation holds along such a curve, in the form

$$P\dot{u}_1 + Q\dot{u}_2 = R\dot{x}, \tag{2.27}$$

where P, Q and R are functions of $\mathbf{u}$, x and y (Exercise 2.7). This expresses the condition that the same linear combination of the right-hand sides of (2.21) and (2.25) must sum to zero.

As an aside we remark that (2.27) can be derived more formally in terms of the *Fredholm Alternative*, which we will encounter in so many guises throughout the rest of the book as to justify our digressing to give a brief description here. The Alternative is motivated by the real linear algebraic equation

$$\mathcal{A}\mathbf{x} = \mathbf{b}, \tag{2.28}$$

for a column vector $\mathbf{x}$, where $\mathcal{A}$ is an $n \times n$ matrix. Suppose that there is a row vector[17] $\mathbf{y}^{\mathsf{T}}$ such that

$$\mathbf{y}^{\mathsf{T}}\mathcal{A} = \mathbf{0}^{\mathsf{T}}, \tag{2.29}$$

i.e. zero is an eigenvalue of $\mathcal{A}^{\mathsf{T}}$ (and hence of $\mathcal{A}$), and $\mathbf{y}^{\mathsf{T}}$ is the corresponding left eigenvector. Then, premultiplying (2.28) by $\mathbf{y}^{\mathsf{T}}$, it is clear that $\mathbf{x}$ can only exist if

$$\mathbf{y}^{\mathsf{T}}\mathbf{b} = 0.$$

Hence the Alternative: either $\mathcal{A}$ is invertible, so that $\mathbf{y}$ is necessarily zero and $\mathbf{x}$ is unique, or $\mathcal{A}$ is not invertible and there are non-zero $\mathbf{y}$, in which case $\mathbf{b}$ must be orthogonal to all such $\mathbf{y}$ if $\mathbf{x}$ is to exist. In the latter case, $\mathbf{x}$ is not unique, because any solution of $\mathcal{A}\mathbf{x} = \mathbf{0}$, i.e. any right eigenvector of $\mathcal{A}$ corresponding to the zero eigenvalue, can be added to it.

[17] Here and henceforth we denote the transpose of a vector or matrix by a superscript T.

Now let us rephrase the Alternative in a way that will be extremely useful later. Forget for the moment that matrices can be transposed, and simply suppose that there is a matrix $\mathcal{A}^*$ such that, for all $\mathbf{z}$ and $\mathbf{w}$,

$$\mathbf{z}^{\mathrm{T}}\mathcal{A}^*\mathbf{w} = \mathbf{w}^{\mathrm{T}}\mathcal{A}\mathbf{z}. \tag{2.30}$$

Now let $\mathbf{w} = \mathbf{y}$. Since $\mathbf{z}$ is arbitrary, we see that $\mathbf{y}$ satisfies (2.29) if and only if $\mathcal{A}^*\mathbf{y} = \mathbf{0}$. The Fredholm Alternative now says that, for existence to be assured, $\mathbf{b}$ must be orthogonal to the vectors annihilated by $\mathcal{A}^*$. Of course, for matrices $\mathcal{A}^*$ is just the transpose of $\mathcal{A}$, but the advantage of (2.30) as a definition of $\mathcal{A}^*$ is that it avoids the use of the transpose, which does not easily generalise to linear partial differential operators. However, (2.30) immediately suggests a formulation of the Alternative that forms the basis of many of the ideas for solving linear partial differential equations in Chapters 4–6.

Conditions (2.27) and (2.26) correspond to the characteristic equations (1.7) in the scalar case but they are now no longer sufficient to determine u_1 and u_2. Indeed (2.27), which represents just one equation for two unknowns u_1 and u_2, is only 'integrable'[18] along the characteristic to give a relation between u_1 and u_2 in special circumstances. Even for a linear problem for which real characteristics can be obtained by integrating (2.26), enabling P, Q and R to be evaluated as functions of t, (2.27) is not sufficient in general to determine u_1 and u_2 unless either $R = 0$ or the ratio Q/P is independent of t. For a nonlinear system in which P, Q and R are functions of u_1 and u_2 but not of x and y, (2.27) is integrable if either $R = 0$ or

$$\frac{\partial}{\partial u_2}\left(\frac{P}{R}\right) = \frac{\partial}{\partial u_1}\left(\frac{Q}{R}\right).$$

Functions of $\mathbf{u}$ which satisfy (2.27) along characteristics are called *Riemann invariants*. Their existence creates a considerable simplification in the structure of the problem and in certain special cases it may lead to solutions in terms of simple functions. It is especially fortunate that these cases include many physically relevant solutions of the gasdynamics systems (2.3)–(2.7). However, the underdeterminacy of (2.27) in general cases explains why partial differential equations cannot usually be reduced to first-order systems of ordinary differential equations. It is only for scalar equations involving only first derivatives that such a reduction is possible in general.

Finally, we note that, as in §1.6, an alternative definition of a characteristic for the system (2.21) is a curve in the (x, y) plane across which $\partial\mathbf{u}/\partial x$ and $\partial\mathbf{u}/\partial y$ may be discontinuous. The generalisation of (1.16) and (1.17) is

$$\mathbf{A}\left[\frac{\partial\mathbf{u}}{\partial x}\right]_-^+ + \mathbf{B}\left[\frac{\partial\mathbf{u}}{\partial y}\right]_-^+ = \mathbf{0} = \dot{x}\left[\frac{\partial\mathbf{u}}{\partial x}\right]_-^+ + \dot{y}\left[\frac{\partial\mathbf{u}}{\partial y}\right]_-^+,$$

so that (2.26) is easily retrieved.

[18]The sense in which (2.27) can be integrated is an interesting question, especially when we consider equations with more independent variables, and we will return to this in §2.4 and Chapter 9.

2.3 The Cauchy–Kowalevski theorem

We now discuss the most important theorem on the existence of solutions to a general first-order quasilinear system

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x}+\mathbf{B}\frac{\partial \mathbf{u}}{\partial y}=\mathbf{c},$$

with n dependent variables and two independent variables. We shall impose Cauchy data on $x = 0$, so it is crucial that this is not a characteristic. Setting $x' = 0$ in the condition (2.24), $\det(x'\mathbf{B} - y'\mathbf{A}) \neq 0$, a necessary and sufficient condition that $x = 0$ is not a characteristic is that $\mathbf{A}$ is invertible, and the system can be then solved for $\partial\mathbf{u}/\partial x$. Furthermore, after multiplying by $\mathbf{A}^{-1}$, we can simultaneously remove the inhomogeneous term $\mathbf{A}^{-1}\mathbf{c}$ and make $\mathbf{A}^{-1}\mathbf{B}$ autonomous (i.e. independent of x and y) by introducing two new dependent variables (the details are given in Exercise 2.5) to give a system of dimension $n + 2$.

We therefore restrict our attention to the homogeneous problem in the autonomous form

$$\frac{\partial \mathbf{u}}{\partial x}=\mathbf{D}(\mathbf{u})\frac{\partial \mathbf{u}}{\partial y}, \tag{2.31}$$

where $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{D}$ is an $n \times n$ matrix.[19] Let us focus on the *Cauchy problem* of finding a solution of (2.31) in $x > 0$ (for definiteness) which satisfies the Cauchy data

$$\mathbf{u}(0,y)=\mathbf{u}_0(y), \tag{2.32}$$

where $\mathbf{u}_0$ is a differentiable function of y. The restriction that the boundary is precisely the y axis does not lose us any generality; if the components of $\mathbf{u}$ were prescribed on some other sufficiently smooth arc, we could change variables to transform that arc into $x = 0$, provided of course that it was nowhere parallel to a characteristic, so that the derivative of $\mathbf{u}$ normal to the curve could be computed at every point.

Now if $\mathbf{u}_0$ is prescribed, direct differentiation gives $\partial\mathbf{u}/\partial y\,(0,y)$ and hence we can find $\partial\mathbf{u}/\,\partial x\,(0,y)$ from (2.31). In a similar way we may compute all the higher derivatives of $\mathbf{u}$ by further differentiation of $\mathbf{u}_0$ and the partial differential equation (2.31), assuming that this differentiation is allowed. The Cauchy–Kowalevski theorem gives conditions under which the resulting double Taylor series expansion about any point of $x = 0$ is convergent and unique. Before outlining the proof we make some remarks to motivate the theorem and emphasise its highly restrictive limitations.

First, we recall that the analogous theorem for an ordinary differential equation is the Cauchy–Picard theorem in which the problem is rewritten as an integral equation and the proof only requires the relatively weak condition that

[19] As suggested in the Introduction, a fairly general first-order scalar equation can be reduced to a quasilinear system, and hence to the form (2.31), provided that it can be solved for the first derivative with respect to x. The same is true for an nth-order equation but, as illustrated in Exercise 2.4, we also see that the dimension of the system obtained may be greater than n. We will see that (2.31) has a great conceptual advantage over the original formulation $\mathbf{A}\,\partial\mathbf{u}/\partial x + \mathbf{B}\,\partial\mathbf{u}/\partial y = \mathbf{c}$ when it comes to the consideration of existence and uniqueness of solutions.

the integrand be a Lipschitz-continuous function of $\mathbf{u}$. Such an integral equation formulation is not generally possible for partial differential equations and we may expect that much stronger conditions have to be imposed on $\mathbf{D}$ and $\mathbf{u}_0$, with correspondingly weaker results about the domain of definition, than the Cauchy–Picard theorem provides. Second, let us consider the innocuous-looking example

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y},$$

for which

$$\mathbf{D} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

with data

$$u(0, y) = f(y), \quad v(0, y) = 0.$$

These are the Cauchy–Riemann equations for $u + \mathrm{i}v$ to be an analytic function of $z = x + \mathrm{i}y$, and hence the solution is

$$u + \mathrm{i}v = f(-\mathrm{i}z)$$

($-\mathrm{i}z = y$ on $x = 0$), provided $f(y)$ is a *real analytic* function of y, namely one that has a Taylor series expansion which converges to f in a neighbourhood of $y = 0$.[20] However, if f is not analytic at a point, e.g. $f(y) = |y|$ and we are near $y = 0$, then there is *no analytic solution* to the Cauchy problem in any neighbourhood of the origin. Equally disastrously, if, for example, $f(y) = \epsilon/(y^2 + \delta^2)$, where $\epsilon, \delta > 0$, then

$$u + \mathrm{i}v = \frac{\epsilon}{\delta^2 - z^2}$$

and

$$u = \frac{\epsilon}{2\delta}\left(\frac{\delta - x}{y^2 + (x - \delta)^2} + \frac{\delta + x}{y^2 + (x + \delta)^2}\right).$$

Thus, no matter how small ϵ is, u fails to exist by becoming infinite at $y = 0$, $x = \delta$, so that the boundary of the domain of definition of u may be arbitrarily close to $y = 0$. Thus, the Cauchy–Riemann system provides a striking illustration of the failure of well-posedness as defined after (2.24).

This example motivates the emphasis on the words *analytic* and *local* in the following statement of the *Cauchy–Kowalevski theorem.*

Cauchy–Kowalevski theorem *If* $\mathbf{u}_0(y)$ *is* analytic *at* $y = 0$ *and* $\mathbf{D}(\mathbf{u})$ *is* analytic *at* $\mathbf{u} = \mathbf{u}_0(0)$, *then the Cauchy problem (2.31) and (2.32) has a* unique analytic *solution*[21] locally *near* $x = y = 0$.

[20]This result is interesting because it shows that the solution u, v of the Cauchy–Riemann system can be thought of as the *analytic continuation* of the function $f(y)$ off the y axis. What we are about to say is a restatement of warnings to be found in some books on complex variable theory that analytic continuation is a dangerous process.

[21]Note that this does not exclude the possibility that *non-analytic* solutions may exist and be non-unique; this question is addressed by Holmgren's theorem.

The word *locally* is particularly important and prevents the theorem from saying anything about well-posedness.

We will only sketch the proof, to avoid cumbersome algebra; more details can be found in [12]. We begin by considering the scalar case for which (2.31) and (2.32) become

$$\frac{\partial u}{\partial x} = d(u)\frac{\partial u}{\partial y}, \quad u(0,y) = u_0(y), \tag{2.33}$$

and let us ignore for the moment the fact that (2.33) can be solved explicitly using the methods of Chapter 1. Since d and u_0 are analytic we can write down convergent Taylor expansions

$$d(u) = \sum_{n=0}^{\infty} d_n u^n, \quad u_0(y) = \sum_{n=1}^{\infty} a_n y^n, \tag{2.34}$$

where we have, if necessary, subtracted the constant $u_0(0)$ from u. Our aim is to show that u itself has a convergent Taylor expansion, so we seek a solution

$$u = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} c_{mn} x^m y^n. \tag{2.35}$$

Substituting in (2.33), we obtain c_{mn} as a uniquely-determined polynomial in $\{d_n\}$ and $\{a_n\}$, moreover one with positive integer coefficients, a fact which is central to the proof.

Now let the radii of convergence of the series in (2.34) be R_d and R_a, respectively. Take two fixed numbers $\rho_d < R_d$ and $\rho_a < R_a$. Because $\sum d_n \rho_d^n$ converges, the supremum over n of $|d_n|\rho_d^n$ clearly exists, and so likewise does the supremum of $|a_n|\rho_a^n$. We denote these numbers by M_d and M_a (they may be quite large). Now consider the comparison function U that satisfies

$$\frac{\partial U}{\partial x} = D(U)\frac{\partial U}{\partial y}, \quad U(0,y) = U_0(y), \tag{2.36}$$

where the Taylor series

$$D(U) = \sum_{n=0}^{\infty} D_n U^n = \sum_{n=0}^{\infty} \frac{M_d}{\rho_d^n} U^n,$$

$$U_0 = \sum_{n=1}^{\infty} A_n y^n = \sum_{n=1}^{\infty} \frac{M_a}{\rho_a^n} y^n$$

have positive coefficients. If this function U has the Taylor expansion

$$U = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} C_{mn} x^m y^n, \tag{2.37}$$

we see that the relation between C_{mn} and D_n, A_n is the same as that between c_{mn} and d_n, a_n. Because this relation is a polynomial with positive coefficients, it is easy to see by the triangle inequality that

$$|c_{mn}| < |C_{mn}| = C_{mn}.$$

However, (2.36) is

$$\frac{\partial U}{\partial x} = \frac{M_d}{1 - U/\rho_d}\frac{\partial U}{\partial y}, \quad U(0,y) = \frac{M_a y}{\rho_a(1 - y/\rho_a)}, \tag{2.38}$$

which can be solved explicitly as in §1.3 and expanded in the form (2.37). This gives the convergence of (2.37) and hence, by comparison, the convergence of (2.35) to the unique analytic solution of the Cauchy problem. Now, of course, the problem (2.38) does not represent any real simplification over the problem (2.33), and in any event both can be solved as in Chapter 1. However, the argument above carries over almost unchanged when (2.33) is a vector equation for $\mathbf{u}$, in which case it is *not* solvable by the methods of Chapter 1. In this case (2.38) also becomes a vector equation for $\mathbf{U}$, but it transpires that all the components of $\mathbf{U}$ are equal so that $\mathbf{U}$ can still be found explicitly as the solution of a scalar equation.

Note that the region in which the theorem guarantees existence of the solution is restricted both by the size of ρ_a and ρ_d and by the possibility of blow-up of U; the latter may impose quite a severe restriction when M_a and M_d are large, and this emphasises the local nature of the theorem (see Exercise 2.6).

2.4 Hyperbolicity

We cannot overemphasise the restrictions on $\mathbf{D}$ and $\mathbf{u}$ that the Cauchy–Kowalevski theorem demands, nor the requirement that the partial differential equation can be written in the form $\partial\mathbf{u}/\partial x = \mathbf{D}(\mathbf{u})\,\partial\mathbf{u}/\partial y$. However, even when all these restrictions are met, because the theorem uses Taylor series expansions, it only tells us about *local* existence and uniqueness and the Cauchy–Riemann system demonstrates how local the existence result can be.

There are now two ways in which we can try to develop a more *global* theory for those partial differential equations whose solutions can be expected to exist away from any curve on which boundary conditions are posed. We could either look for solutions of $\partial\mathbf{u}/\partial x = \mathbf{D}(\mathbf{u})\,\partial\mathbf{u}/\partial y$ which satisfy boundary conditions different from, and less restrictive than, the Cauchy data (2.32) (for example, we could prescribe fewer than n components of $\mathbf{u}$ on the boundary), or we could ask for further restrictions on $\mathbf{D}$ whereby the Cauchy problem is well posed and hence has a solution which is not restricted to some very small neighbourhood of the y axis. We recall that in the first-order scalar equations of Chapter 1 no such restriction was needed on the coefficient $\mathbf{D}$.

The former approach leads us to the initial and boundary value problems for elliptic and parabolic equations to be discussed in Chapters 5 and 6. The latter idea leads to the concept of *hyperbolicity* and *hyperbolic equations*, which will be described in detail in Chapter 4 for scalar second-order equations. However, we are already in a position to describe an intuitive framework within which to answer the question as to how $\mathbf{D}(\mathbf{u})$ should be restricted for the Cauchy problem for a first-order system to be well posed.

2.4.1 Two-by-two systems

We saw in §2.2 that, in certain cases, Riemann invariants exist for a system of dimension two and hence a certain known function is constant along any characteristic. Suppose now that the partial differential equation $\mathbf{A}\,\partial\mathbf{u}/\partial x+\mathbf{B}\,\partial\mathbf{u}/\partial y=\mathbf{c}$ is in fact such that (2.26) yields two real distinct characteristics $C_{1,2}$, and that the consistency condition (2.27) can be integrated for both the Riemann invariants. This occurs, for example, if $\mathbf{A}$ and $\mathbf{B}$ are constant and (2.26) has real distinct roots for $\dot{y}/\dot{x}$. In this fortunate circumstance, if we prescribe $\mathbf{u}$ on a boundary Γ which is nowhere tangent to $C_{1,2}$, we can find $\mathbf{u}(P)$ by merely solving the pair of simultaneous *algebraic* equations given by the Riemann invariants at P (see Fig. 2.6(a)). We can then immediately assert that the solution at P exists, is unique and depends continuously on the data until or unless some pathology develops in the solution of these algebraic equations. Thus the Cauchy problem is well posed.

In more general cases, we could imagine taking a sequence of points P_i close to Γ and iterating as indicated in Fig. 2.6(b), regarding $\mathbf{A}$ and $\mathbf{B}$ as constants in each iteration. Hence, even when equations (2.26) and (2.27) do not have integrals, we conjecture that, for a quasilinear first-order system of dimension two, the Cauchy problem is well posed provided that (2.26) has two real distinct roots at each point $(x, y, \mathbf{u})$ of interest, that is, provided two characteristic directions exist at each point. This provides the motivation for defining a *hyperbolic system* of dimension two as one in which (2.26) has two real distinct roots for $\dot{y}/\dot{x}$. In general, these roots depend on the solution $\mathbf{u}$; even in the semilinear case when $\mathbf{A}$ and $\mathbf{B}$ depend only on x and y, a system may be hyperbolic in only part of a given domain, so that a problem of *mixed type* occurs. For the scalar equation $a\,\partial u/\partial x+b\,\partial u/\partial y=c$ the problem is necessarily hyperbolic, and for the two-by-two vector case (2.21) we distinguish the three possibilities, namely *elliptic*, *parabolic* and *hyperbolic*, depending on whether there are no, one or two real characteristics, respectively. This observation is the basis of the more detailed discussion of Chapter 3, where

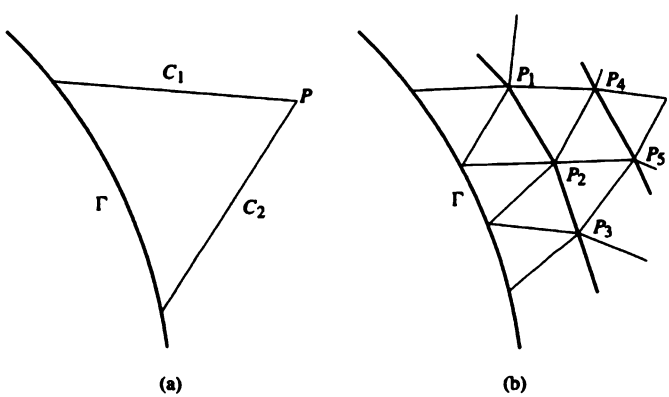


Fig. 2.6 Solution by characteristics: (a) $\mathbf{A}$ and $\mathbf{B}$ constant; (b) approximate solution when $\mathbf{A}$ and $\mathbf{B}$ are functions of $\mathbf{u}$.

we will see that the parabolic case is especially delicate. Indeed, the case where both $\mathbf{A}$ and $\mathbf{B}$ are the identity matrix is parabolic under this classification, even though the system comprises two scalar hyperbolic equations.

2.4.2 Systems of dimension n

With n dependent variables, (2.26) shows that at most n characteristic directions exist at a point. Guided by a motivational argument similar to that which led to Fig. 2.6, the system is said to be *hyperbolic* at that point if (2.26) has n distinct real roots for $\lambda = \dot{y}/\dot{x}$.

In order to consider the possibility of Riemann invariants for systems of dimension n, we define the left eigenvector $\boldsymbol{\ell}$ corresponding to a root $\lambda = \dot{y}/\dot{x}$, so that

$$\boldsymbol{\ell}^{\mathsf{T}}(\mathbf{B} - \lambda\mathbf{A}) = \mathbf{0}^{\mathsf{T}}.$$

Premultiplying the partial differential equation by $\boldsymbol{\ell}^{\mathsf{T}}$, we obtain

$$\boldsymbol{\ell}^{\mathsf{T}}\mathbf{A}\left(\frac{\partial \mathbf{u}}{\partial x} + \lambda\frac{\partial \mathbf{u}}{\partial y}\right) = \boldsymbol{\ell}^{\mathsf{T}}\mathbf{c}.$$

If t is a scalar variable parametrising the characteristic, the generalisation of (2.27) is

$$\boldsymbol{\ell}^{\mathsf{T}}\mathbf{A}\dot{\mathbf{u}} = \boldsymbol{\ell}^{\mathsf{T}}\mathbf{c}\,\dot{x}. \tag{2.39}$$

Again, it is just one ordinary differential relation for the variation of a combination of the n components of $\mathbf{u}$ along the characteristic. As remarked earlier, only when $n = 1$ is it a first-order ordinary differential equation for u which, in the linear case, reduces to a quadrature. In general, it cannot be integrated except for constant coefficient equations, when, along the characteristic,

$$\boldsymbol{\ell}^{\mathsf{T}}\mathbf{A}\mathbf{u} = \boldsymbol{\ell}^{\mathsf{T}}\int^{x}\mathbf{c}\,\mathrm{d}x'. \tag{2.40}$$

The unlikelihood of integrating (2.39), even in the case $\mathbf{c} = \mathbf{0}$, can be seen at once when $n = 3$, because it is well known that the *Pfaffian*

$$\mathbf{P}\cdot\mathrm{d}\mathbf{u} = P_1(\mathbf{u})\,\mathrm{d}u_1 + P_2(\mathbf{u})\,\mathrm{d}u_2 + P_3(\mathbf{u})\,\mathrm{d}u_3$$

is proportional to a total differential $\mathrm{d}\phi(\mathbf{u})$ if and only if

$$\mathbf{P}\cdot\nabla\wedge\mathbf{P} = 0,$$

the curl being taken with respect to $\mathbf{u}$. Thus, even autonomous equations with $\mathbf{c} = \mathbf{0}$ rarely have Riemann invariants as soon as $n \geqslant 3$.

If n Riemann invariants do exist for a hyperbolic system, then on each characteristic there is one algebraic relation between the n components of $\mathbf{u}$. The existence and uniqueness of $\mathbf{u}$ are thus assured in some neighbourhood of a boundary on which $\mathbf{u}$ is prescribed, assuming the boundary is nowhere parallel to a characteristic. However, we must be careful with the definition of this neighbourhood

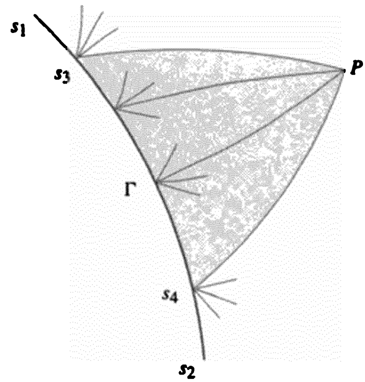


Fig. 2.7 Domain of dependence of P (shaded).

when $n > 2$; for example, if boundary data $\mathbf{u} = \mathbf{u}_0(s)$ is only given on a segment $s_1 \leqslant s \leqslant s_2$ of the initial curve, then clearly the domain of existence is restricted to a region such that all n characteristics at each point P of the region intersect the given segment of the boundary, as in Fig. 2.7 for the case $n = 4$.

If P is the intersection on one side of Γ of the extreme characteristics through $s = s_3$ and $s = s_4$, the domain enclosed by the boundary and these extreme characteristics is called the *domain of dependence* of P. We have only indicated this domain on one side of Γ in Fig. 2.7 because, for most hyperbolic problems that arise as models of physical phenomena with two independent variables, one of these variables is 'time-like'; that is, solutions are only required for which this variable increases away from the 'initial' data on the boundary.[22] The solution at P depends on the boundary data given within its domain of dependence and is independent of boundary data given on sections of the boundary outside it. Equally, a given point Q on the initial curve has a *region of influence* defined by the extreme characteristics through it, as in Fig. 2.8; that is, a change in the boundary data at Q would change the solution everywhere in its region of influence, and only there.

The generalisation of Fig. 2.6 to cases where $n > 2$ suggests that solutions may be constructed numerically by approximating the characteristic curves by straight lines and approximating the differential relations (2.39) holding along characteristics by algebraic relations. Thus Fig. 2.9 shows how the solution at P may be obtained from a knowledge of $\mathbf{u}$ at four different points on the initial curve in the case $n = 4$. By varying P, data on a new initial curve is obtained and the process repeated. This procedure clearly fails if two characteristics become parallel, in which case the problem ceases to be hyperbolic, or if two characteristics of the same family intersect, as happened in §1.4 for the case $n = 1$.

[22]We will see in §2.6 that hyperbolic systems with more than two independent variables allow a much more precise definition to be made of the term time-like.

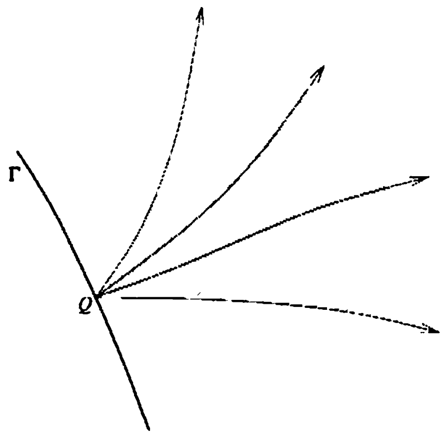


Fig. 2.8 Region of influence of Q (shaded).

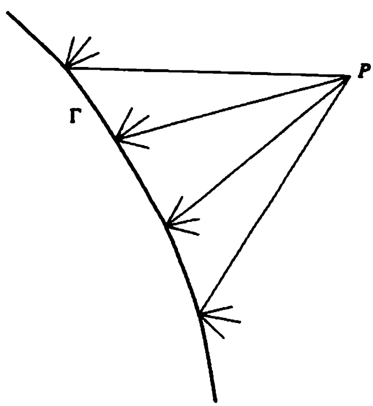


Fig. 2.9 The solution at P using linear approximations to the characteristics.

If Riemann invariants exist, some explicit results may be obtained in closed form. The following examples (which refer to the models introduced in §2.1) and remarks illustrate the kind of information that may be obtained.

2.4.3 Examples

Shallow water theory

In this example, described by (2.1) and (2.2), we identify (x, y) in our general discussion above with (t, x), and we find

$$\lambda = u \pm s, \quad \text{where } s = \sqrt{gh},$$

and the Riemann invariants are $u \pm 2s$, respectively. The system is always hyperbolic unless the river is dry, but the nonlinearity means that the characteristics can often intersect, corresponding to the formation of a 'bore' or 'hydraulic jump'. In §4.8.1 we will see how to use these Riemann invariants to generate some useful explicit solutions to (2.1) and (2.2).

This example illustrates the point that the existence of Riemann invariants is by no means a necessary condition for a first-order system to have explicit solutions, because we cannot usually locate the characteristics explicitly. Among the many clever 'tricks' for guessing explicit solutions, one possibility to keep in mind for autonomous systems such as (2.1) and (2.2) is that of exchanging the roles of dependent and independent variables. We see that, if we attempt the *hodograph transformation*

$$x = X(u,h), \quad t = T(u,h), \tag{2.41}$$

so that, by the chain rule,

$$\frac{\partial u}{\partial t} = \frac{1}{\Delta}\frac{\partial X}{\partial h},$$

etc., with the Jacobian

$$\Delta = \frac{\partial T}{\partial u}\frac{\partial X}{\partial h} - \frac{\partial T}{\partial h}\frac{\partial X}{\partial u}$$

assumed non-zero, then we obtain the *linear* system[23]

$$\frac{\partial X}{\partial h} - u\frac{\partial T}{\partial h} + \frac{\partial T}{\partial u} = 0,$$
$$\frac{\partial X}{\partial u} - u\frac{\partial T}{\partial u} + h\frac{\partial T}{\partial h} = 0,$$

where we have set $g = 1$ for simplicity. This of course only works because there are no undifferentiated terms in the original equations (2.1) and (2.2), but it has an interesting geometrical interpretation which will be discussed further in Chapter 4.

Unsteady one-dimensional gas dynamics

Rearranging (2.4) (with the help of (2.3)) as

$$\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + \gamma p\frac{\partial \rho}{\partial x} = 0,$$

and again identifying (t,x) with (x,y), we find, after some calculations (Exercise 2.10), that the system is always hyperbolic with

$$\lambda = u \quad \text{or} \quad \lambda = u \pm \sqrt{\frac{\gamma p}{\rho}}$$

as the slopes of the characteristics, on which

$$\gamma p\,\mathrm{d}\rho - \rho\,\mathrm{d}p = 0, \quad \mp\sqrt{\gamma p\rho}\,\mathrm{d}u + \mathrm{d}p = 0,$$

respectively. Hence p/ρ^γ is constant on $\mathrm{d}x/\mathrm{d}t = u$ but, in general, Riemann invariants do not exist on all three families of characteristics.

[23] See [30] for a discussion of the corresponding system in gas dynamics.

Steady two-dimensional gas dynamics

Our analysis in §2.2 of the linearised system (2.8)–(2.10) has already revealed the existence of three real distinct characteristics when the basic flow is supersonic, and it is only under these circumstances that the system is hyperbolic. A slightly different situation arises for the full (unlinearised) system (2.5)–(2.7) where, after more lengthy algebraic manipulations (Exercise 2.11), we find

$$\lambda = \frac{v}{u} \text{ twice} \quad \text{or} \quad \lambda = \tan\left(\tan^{-1}\left(\frac{v}{u}\right) \pm \sin^{-1}\sqrt{\frac{\gamma p}{\rho(u^2+v^2)}}\right).$$

Now, there are four real roots when $u^2 + v^2 > \gamma p/\rho$, in which case the flow is again said to be supersonic. However, the system is not strictly hyperbolic even when the flow is supersonic because two of the eigenvalues, namely $\lambda = v/u$, coincide; in practice, however, the solutions usually behave as if the system were hyperbolic.

In both of the examples above, neighbouring members of the second or third families of characteristics found above can intersect in general to form gasdynamic shock waves. Much more will be said about these shock waves in the next section.

Flow of granular materials

The model (2.13) with $h = 1$, together with the relation (2.16) between τ, σ_x and σ_y, can, after a long calculation (started in Exercise 2.3, where ψ is defined, and continued in Exercise 2.14), be shown to be always hyperbolic with

$$\lambda = -\cot\left(\psi \pm \left(\frac{\pi}{4} - \frac{\phi}{2}\right)\right),$$

and the Riemann invariants are

$$\pm\cot\phi \log\left(-\frac{\sigma_x + \sigma_y}{2}\right) + 2\psi,$$

respectively. We remark that these characteristics are *not* the slip planes that were needed to formulate the model.

Fluidised and packed beds

The two linear models (2.17) and (2.18), and (2.19) and (2.20) are clearly always hyperbolic; the characteristics of the latter are given by

$$\frac{dx}{dt} = v_a, v_b,$$

while for the former one of these characteristic 'speeds' is zero. However, the presence of the undifferentiated terms means that there are no explicit Riemann invariants unless k or h is zero. Fortunately, the linearity of these models enables us to circumvent this difficulty, as we shall see in Chapter 4.

As an aside, suppose that $k = 0$ in (2.19) and (2.20) and we were attempting to solve the Cauchy problem in which c_a and c_b, which are now Riemann invariants, were prescribed on the non-characteristic curve $t = 0$. Now suppose further that we were interested in modelling a fluidised bed of finite depth, and we were unsure (as we often are) about the boundary conditions that should be applied at the top and bottom of the bed. One thing we can be sure of at once is that we should *not* impose values of either c_a or c_b at the top of the bed, towards which the gas is moving with velocity v_a or v_b. If such values were prescribed, they would almost certainly clash with the values of the Riemann invariants that follow from the Cauchy data.

Optical fibre model

A transformation slightly less obvious than the hodograph transformation (2.41) can be applied to the optical fibre model (2.11) and (2.12). For a given T, this system is hyperbolic (unlike most models of slow viscous flow), with characteristics $\mathrm{d}x/\mathrm{d}t = u$, the local fluid velocity, and $t =$ constant, corresponding to infinite propagation speed in the incompressible glass. However, the presence of T means that there are no explicit Riemann invariants. Nonetheless, if we write the 'partial' hodograph transformation

$$x = x(A, t), \quad u = u(A, t),$$

and rescale this so that $\tau = \int_0^t T(t')\,\mathrm{d}t'$, we find (Exercise 2.15) that x satisfies the linear equation

$$\frac{\partial^2 x}{\partial A^2} - c\frac{\partial^2 x}{\partial A \partial \tau} = -\frac{1}{A}\frac{\partial x}{\partial A},$$

which can be solved explicitly for $\partial x/\partial A$. This enables us to show that the characteristics never intersect.

The whole question of finding whether any given quasilinear system has enough symmetry to be able to be integrated explicitly is intimately connected with the theory of groups that depend on a continuously-varying parameter. We will return to this on several later occasions but, for now, the all-too-frequent occurrence of shock waves as a result of the intersection of characteristics of hyperbolic systems demands immediate attention.

*2.5 Weak solutions and shock waves

We have shown at the end of §2.2 that, for a hyperbolic system, the first derivatives of $\mathbf{u}$ can be discontinuous across a characteristic, and in Chapter 1 we have given several examples where characteristics may intersect to cause discontinuities in $\mathbf{u}$ itself. Hence, following the discussion of weak solutions for the scalar equations of Chapter 1, we examine the possibility of jump discontinuities in $\mathbf{u}$ across a curve Γ for the system $\mathbf{A}\,\partial\mathbf{u}/\partial x + \mathbf{B}\,\partial\mathbf{u}/\partial y = \mathbf{c}$. Because Green's theorem plays a crucial role, we only really have any hope of generalising the theory presented in §1.7 for a hyperbolic system when it is in conservation form. Hence we consider the problem of defining a weak solution of

$$\frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial y} = \mathbf{c}, \tag{2.42}$$

where $\mathbf{v}$ and $\mathbf{c}$ are functions of $\mathbf{u}$, x and y; for simplicity, initial data $\mathbf{u} = \mathbf{u}_0(y)$ is given on $x = 0$ and the domain of interest is $x > 0$. For (2.42) to be hyperbolic, the matrix $(\partial v_i/\partial u_j)$ must have n distinct real eigenvalues. If we now multiply (2.42) by a test function $\psi(x, y)$, differentiable everywhere in $x > 0$ and such that $\psi \to 0$ sufficiently rapidly as x or $y \to \infty$, then, using Green's theorem on the half plane $x > 0$, we obtain

$$\iint_{x>0} \left(\mathbf{u}\frac{\partial \psi}{\partial x} + \mathbf{v}\frac{\partial \psi}{\partial y} + \psi \mathbf{c} \right) \mathrm{d}x\, \mathrm{d}y = \int_{-\infty}^{\infty} \psi(0, y)\mathbf{u}_0(y)\, \mathrm{d}y. \tag{2.43}$$

The statement (2.43) generalises (1.25), and our restrictions on the behaviour of ψ have left us with a convenient right-hand side. The equation makes sense even if $\mathbf{u}$ is discontinuous, and hence we define a weak solution of (2.42) to be a function $\mathbf{u}$ satisfying (2.43) for all suitable test functions ψ. If the weak solution $\mathbf{u}$ is continuous then clearly, retracing our steps, (2.43) implies (2.42) and a classical solution is a weak solution. A weak solution need not, however, be continuous; if it is discontinuous across a curve Γ then, carrying over the argument leading to (1.27) and applying Green's theorem to (2.43) for the domain $x > 0$ excluding Γ, we obtain

$$\int_\Gamma [\psi \mathbf{u}]_-^+ \,\mathrm{d}y = \int_\Gamma [\psi \mathbf{v}]_-^+ \,\mathrm{d}x. \tag{2.44}$$

Since ψ is continuous across Γ and (2.44) holds for all such ψ, then, necessarily,

$$[\mathbf{u}]_-^+ \frac{\mathrm{d}y}{\mathrm{d}x} = [\mathbf{v}]_-^+, \tag{2.45}$$

which is the jump or Rankine–Hugoniot condition for a general hyperbolic system of conservation laws.

For a semilinear problem, $\mathbf{v}$ is equal to $\mathbf{D}\mathbf{u}$, where $\mathbf{D}$ is a matrix which does not depend on $\mathbf{u}$, and hence $\mathrm{d}y/\mathrm{d}x$ is an eigenvalue of $\mathbf{D}$. In this case discontinuities in $\mathbf{u}$ can occur across any of the characteristics. In general, however, they lie on curves called ***shocks***, which are definitely ***not*** characteristics.

Of course, the approach above inherits all the non-uniqueness properties that were encountered in Chapter 1 for the case $n = 1$. We will now describe two ways in which this non-uniqueness might be resolved.

2.5.1 Causality

We begin by remarking that we can construct weak solutions by piecing together classical solutions using the jump condition (2.45). At a point P on the shock there are n characteristic directions on each side of it, as in Fig. 2.10 for the case $n = 2$. These $2n$ characteristics through P form two classes, those that intersect $x = 0$ and those that do not.

On each characteristic which intersects $x = 0$, some information propagates about the boundary data $u_0(y)$; in the special case when a Riemann invariant exists on the characteristics, this takes the form of a prescribed algebraic relation between the components of $\mathbf{u}$ as we approach the shock (from the left in Fig. 2.10). For the shock strength $[\mathbf{u}]$ and direction to be defined uniquely it is necessary that

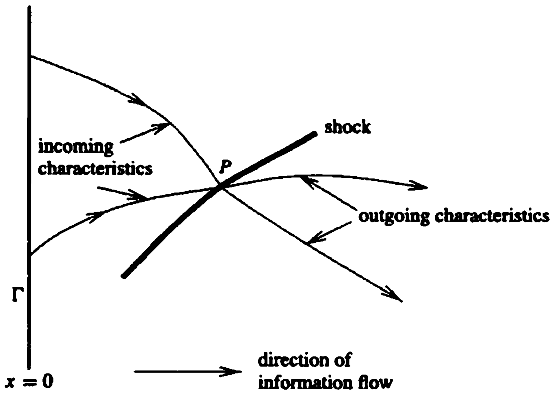


Fig. 2.10 Incoming and outgoing characteristics near a shock; causality.

the correct number of relations be available to determine the values of $\mathbf{u}$ just to the right of the shock. Thus, if k characteristics at P are 'outgoing', that is they do not intersect $x = 0$, then there are $2n - k$ relations to determine the $2n$ unknown values of the components of $\mathbf{u}$ on either side of the shock, together with n Rankine–Hugoniot conditions (2.45) involving the unknown shock slope $\mathrm{d}y/\mathrm{d}x$. That totals $3n - k$ relations for $2n + 1$ unknowns and hence $k = n - 1$.

This simple-minded argument forms the basis for the method of *causality* for uniquely defining the weak solution: any physically admissible shock must be such that $n - 1$ characteristics through any point on it are outgoing.[24]

To illustrate the use of this causality method we consider two examples with $n = 1$ and $n = 3$, whose classical solutions have already been discussed.

Example 2.2 (*Equation (1.14) revisited*) We have already encountered the equation

$$\frac{\partial u}{\partial x} + \frac{\partial}{\partial y}\left(\frac{1}{2}u^2\right) = 0 \tag{2.46}$$

in §1.5, and it can be shown that in certain circumstances it provides an approximation to the system (2.3) and (2.4) describing one-dimensional gas dynamics, with x and y being time and space, respectively. In this context, an experimental device known as a ***shock tube*** can be modelled by the initial conditions

$$u_0 = \begin{cases} 0, & y < 0, \\ 1, & y > 0, \end{cases} \tag{2.47}$$

for which two solutions were found in §1.7, one of which is continuous. The discontinuous solution is

$$u = \begin{cases} 0, & y < x/2, \\ 1, & y > x/2, \end{cases}$$

[24]We will describe this situation in more detail in §7.2.4 for a scalar equation.

and the characteristic directions on the shock $y = x/2$ have slope zero in $y < x/2$ and unity in $y > x/2$, as in Fig. 2.11. They are therefore both outgoing so that $k = 2 > n - 1$; the causality condition rejects this solution (we could indeed construct many such inadmissible shocks). In fact, this argument applies to all the weak solutions that we could find that satisfy the Rankine–Hugoniot relations with a shock of non-zero strength. Hence there are no causal discontinuous solutions and the continuous solution

$$u = \begin{cases} 0, & y < 0, \\ y/x, & 0 \leqslant y \leqslant x, \\ 1, & x < y, \end{cases}$$

which is automatically causal, is the only acceptable solution to the Cauchy problem.

Now let us look at the case

$$u_0 = \begin{cases} 0, & y > 0, \\ 1, & y < 0. \end{cases} \tag{2.48}$$

A solution, indeed the *only* one satisfying the Rankine–Hugoniot relations, is

$$u = \begin{cases} 0, & y > x/2, \\ 1, & y < x/2; \end{cases}$$

the characteristic directions on the shock have slope zero in $y > x/2$ and unity in $y < x/2$, as in Fig. 2.12. They are therefore both incoming, that is they both intersect $x = 0$, and hence $k = 0$. Thus $k = n - 1$ and this is indeed the unique appropriate causal solution.

This example paves the way for a more complete discussion of gasdynamic shock waves. The models (2.3)–(2.7) are already in conservation form and the jump conditions determined from (2.45) represent the well-known Rankine–Hugoniot

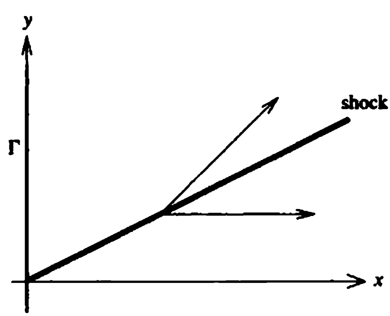


Fig. 2.11 Solution of (2.46) and (2.47) with too many outgoing characteristics.

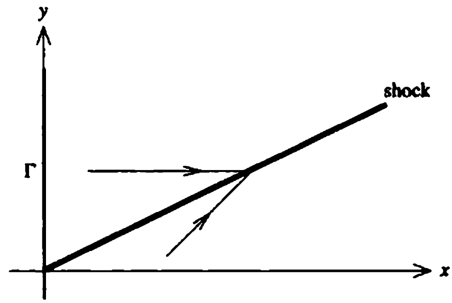


Fig. 2.12 Causal solution of (2.46) and (2.48).

relations for gasdynamic shocks. For a one-dimensional shock moving with speed V the result of applying (2.45) to (2.3) and (2.4) can be manipulated to yield

$$[\rho(V-u)]_-^+ = 0, \qquad [p+\rho(V-u)^2]_-^+ = 0,$$
$$\left[\frac{\gamma p}{(\gamma-1)\rho} + \frac{1}{2}(V-u)^2\right]_-^+ = 0, \tag{2.49}$$

corresponding to conservation of mass, momentum and energy, respectively (see Exercise 2.17). For the existence of a shock, with $[p]_-^+ \neq 0$, separating two regions in the (x,t) plane in which the flow variables are constant, it can be shown further that the flow must pass through a 'compressive shock', across which p and ρ increase, if it is to be modelled by a causal weak solution. Figure 2.13 shows the (x,t) plane for (2.3) and (2.4) corresponding to a piston $x = Vt$ driving a gas in $x > Vt$. For $V > 0$ we have $n = 3$ and the causal weak solution has $k = 2$. However, this value of k cannot be attained when $V < 0$, and thus no 'expansion shock' appears in this case.

A similar argument can be applied to the two-dimensional system (2.5)–(2.7) and, after lengthy algebra, it emerges that two possible Rankine–Hugoniot shocks can 'bend' a prescribed supersonic stream around a corner, as in Fig. 2.14. Now $n = 4$ (although one characteristic is double) and the appropriate weak solution has $k = 3$; the shock with the smaller slope is the physically acceptable solution, as is readily observed in wind tunnels.[25]

2.5.2 Viscosity and entropy

Causality is just one kind of mathematical filter that can be introduced in an effort to try to render weak solutions of conservation laws unique. Two other methods

[25] We remark that other 'shocks' that satisfy the Rankine–Hugoniot relation are possible. For example, it is possible to satisfy (2.49) with $[p]_-^+ = 0$, $[u]_-^+ = 0$, $u = U$ and $[\rho]_-^+ \neq 0$, and this is called a *contact discontinuity*. For the two-dimensional system (2.5)–(2.7) the analogous shock has a jump in density and tangential velocity and it is called a *vortex sheet*, a concept to which we will return in Chapter 7. Both these 'shocks' are degenerate in that they coincide with one of the characteristics.

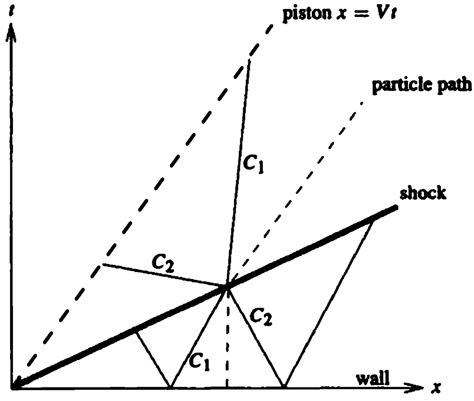


Fig. 2.13 Characteristics for a shock generated by a piston (the particle paths are also characteristics).

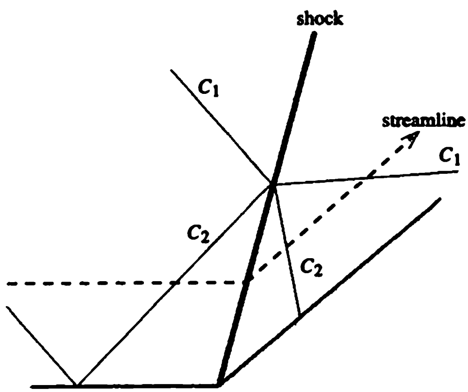


Fig. 2.14 Characteristics for a supersonic gas stream in a corner (the streamlines are also characteristics).

devoted to the same goal are associated with the names of 'viscosity' and 'entropy' but in fact all three methods are closely related to each other.

Viscosity methods can be most easily illustrated by referring back to the example above, (1.14), thinking of it again in the context of gas dynamics. It can be shown by modelling arguments, to which we will refer in more detail in Chapter 6, that the effect of a small amount of viscosity ϵ in the one-dimensional flow model (2.3) and (2.4) has the effect of replacing (1.14) by

$$\frac{\partial u}{\partial x} + \frac{\partial}{\partial y}\left(\frac{1}{2}u^2\right) = \epsilon\frac{\partial^2 u}{\partial y^2}, \tag{2.50}$$

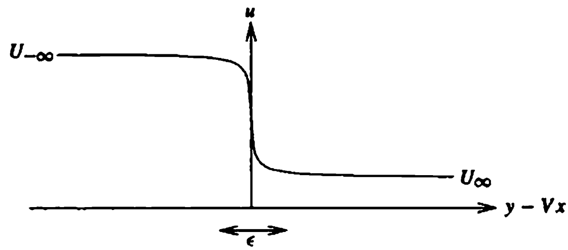


Fig. 2.15 Smoothing of a shock by viscosity.

at least as far as relatively small disturbances are concerned.[26] As shown in Exercise 2.18, this equation, which is known as *Burgers' equation*, has solutions that are travelling waves. These are such that $u = U\big((y-Vx)/\epsilon\big)$, $-\infty < y-Vx < +\infty$, as long as

$$V = \left[\frac{1}{2}U^2\right]_{-\infty}^{\infty} \Big/ [U]_{-\infty}^{\infty};$$

this is just the condition that the Rankine–Hugoniot conditions are satisfied by the asymptotic values of u far 'upstream' and 'downstream' of the shock wave. The shock has been smeared out over a region in which $y - Vx$ is of order ϵ by the presence of the viscosity, as in Fig. 2.15.

The introduction of viscosity not only retrieves the Rankine–Hugoniot conditions for us, but also still more information is contained in (2.50). The analysis of Exercise 2.18 shows that, in a travelling wave, U must be a decreasing function of $y - Vx$. We will see in Chapter 6 that this has the physical interpretation that the wave can only be compressive, and hence we have excluded 'expansion' shocks, as likewise did the causality condition for (2.50) with $\epsilon = 0$.

A further idea that is equivalent to causality in this case is that, if (1.14) is again rewritten as

$$\frac{\partial u}{\partial x} + \frac{\partial}{\partial y}\left(v(u)\right) = 0,$$

then the convex function $v = \frac{1}{2}u^2$, known as the *entropy*, has to *increase* as a result of the passage of the shock, irrespective of the sign of the shock velocity. Thus an analogy can be drawn between causal shocks, viscous shocks and shocks across which a convex function increases. The latter interpretation suggests an analogy with the concept of entropy that arises in statistical mechanics.

In summary, the theory of weak solutions of hyperbolic conservation laws is incomplete without the incorporation of extra information. When this information is introduced in the form of causality, viscosity or entropy arguments, a unique

[26] Note that the introduction of such a viscosity term into the right-hand side of a conservation law, which is the result of a modelling assumption, has different consequences depending on the way in which the law is written; the solution of (2.50) would be quite different if the left-hand side were $\partial/\partial x\left(\frac{1}{2}u^2\right) + \partial/\partial y\left(\frac{1}{3}u^3\right)$; viscosity thus legislates for a particular weak solution, reflecting the modelling assumption that has been made.

weak solution emerges in which the Rankine–Hugoniot conditions are satisfied across any shock waves that are present. The concept of viscosity is a stronger one than those of causality or entropy because it even allows us to predict the Rankine–Hugoniot conditions as well as selecting a unique limiting solution as the viscosity tends to zero.

2.5.3 Other discontinuities

A much more general discussion of discontinuous solutions of partial differential equations will be given in Chapter 7. That chapter is rather lengthy, mostly because of the delicate nature of the subject, and there could be no better example of this than the modelling of gasdynamic shock waves in the presence of chemical reactions which are localised near the discontinuity, which is then often called a ***detonation***. In this case, the right-hand side of (2.4) is replaced by a localised source term, which means that the Rankine–Hugoniot relation for the energy changes from (2.49) to

$$\left[\frac{\gamma p}{(\gamma-1)\rho}+\frac{1}{2}(V-u)^2\right]_-^+=E,$$

where E is the prescribed constant energy per unit mass released during the reaction. When E is zero, we can, as mentioned above, derive non-unique weak solutions in which $[p]_-^+$, $[\rho]_-^+$ and $[u]_-^+$ satisfy the Rankine–Hugoniot relations (2.49). It can be shown that we can rearrange these relations into

$$\tilde{p}=\frac{\gamma+1-(\gamma-1)/\tilde{\rho}}{(\gamma+1)/\tilde{\rho}-(\gamma-1)},$$

where $\tilde{p}$ and $\tilde{\rho}$ are the ***ratios*** of the downstream to the upstream pressures and densities, respectively (upstream referring to gas that has not yet traversed the discontinuity). Thus $(\tilde{p},1/\tilde{\rho})$ lies on a hyperbola called the Chapman–Jouguet (C–J) curve, as shown in Fig. 2.16(a), with the point $(1,1)$ corresponding to the upstream condition. The unique causal weak solution corresponds to $\tilde{p}>1$, $1/\tilde{\rho}<1$, which is why the lower half of the C–J curve is shown dotted. All points on the upper half represent admissible shocks.

However, when E is positive the denominator in the $\tilde{p}$, $\tilde{\rho}$ relation is increased by a positive multiple of E and the new C–J curve is shown relative to the old one in Fig. 2.16(b); the point $(1,1)$ still represents the upstream condition. The middle section of this curve is shown broken and is irrelevant, because it is easy to see from the other Rankine–Hugoniot relations that, when $\tilde{p}>1$ and $1/\tilde{\rho}>1$, the velocity of the gas traversing the discontinuity is in the wrong direction, from downstream to upstream. What is most important, however, is that when $E>0$ we can no longer resort to entropy or causality arguments to render points on the lower half of the C–J curve inadmissible. Compressive detonations certainly exist, corresponding to the upper continuous branch of the curve in Fig. 2.16(b). However, expansive 'deflagrations', corresponding to the lower continuous branch, can also occur, although they are more prone to instability, as people who run explosives laboratories know well.

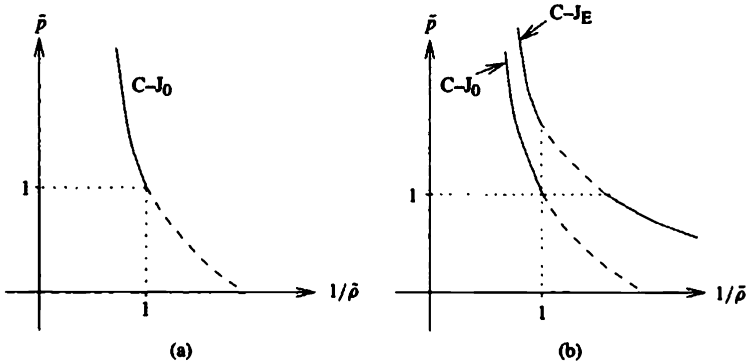


Fig. 2.16 The Chapman–Jouguet curve (a) without energy release, $E = 0$, and (b) with energy release, $E > 0$.

*2.6 Systems with more than two independent variables

It would be nice to be able to end the chapter in the same way as we did in Chapter 1 by saying that the ideas developed thus far can be extended trivially to systems with more than two independent variables. Unfortunately, apart from the Cauchy–Kowalevski theorem, this is so far from being the case that readers with a nervous disposition should perhaps refrain from scanning the following pages and skip directly to Chapter 3.

When we attempt to generalise the ideas of characteristics to equations with more than two independent variables, we immediately run into difficulties with geometric visualisation. This is because, if we consider the general $m \times n$ system

$$\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \mathbf{u}}{\partial x_i} = \mathbf{c}, \qquad \mathbf{u} = (u_j) \quad \text{for } 1 \leqslant j \leqslant n, \tag{2.51}$$

we have to seek the solution to the Cauchy problem by considering initial data for which $\mathbf{u}$ is given on a manifold of dimension $m - 1$. Now if we ask what $(m-1)$-dimensional manifolds are such that the partial derivatives of $\mathbf{u}$ normal to the manifold are not well defined, it is easiest to proceed by denoting such a manifold by, say,

$$\phi(x_1, \ldots, x_m) = 0,$$

and hence regard the manifold as a level set of a family of functions ϕ that are smooth enough that we can change to ϕ as a new independent variable. Then (2.51) simply tells us that

$$\left(\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \phi}{\partial x_i} \right) \frac{\partial}{\partial \phi} \mathbf{u} \tag{2.52}$$

is equal to a linear combination of $\mathbf{c}$ and the derivatives of $\mathbf{u}$ tangential to the surface $\phi = 0$, and is therefore known in terms of the Cauchy data. Hence, if we define

$$Q\left(\frac{\partial\phi}{\partial x_1},\ldots,\frac{\partial\phi}{\partial x_m}\right) = \det\left(\sum_{i=1}^{m}\mathbf{A}_i\frac{\partial\phi}{\partial x_i}\right),$$

we are led to define a characteristic surface $\phi = 0$ as one on which

$$Q\left(\frac{\partial\phi}{\partial x_1},\ldots,\frac{\partial\phi}{\partial x_m}\right) = 0. \tag{2.53}$$

This clearly reduces to the first two of (1.7) and (2.26) when $m = 2$ and $n = 1$, 2, respectively. Equation (2.53) also reduces to (1.32) in the scalar case, and our remarks about the interpretation of (1.32) as a partial differential equation apply equally to (2.53).

In principle, it is easy to generalise the Cauchy–Kowalevski theorem to prove the existence and uniqueness of solutions to the Cauchy problem when the data is given on a non-characteristic manifold, given the usual crucial requirements of locality and analyticity.

On the other hand, when $m > 2$, not only do we have the irksome task of solving a partial differential equation for ϕ, but also it is more difficult to visualise what (2.53) means for the surface $\phi = 0$. However, bearing in mind that the normal to the surface is always $(\partial\phi/\partial x_i) = (\xi_i)$, say, we can see that (2.53), being a homogeneous expression of degree n in these quantities, states that at any point P of $\mathbb{R}^m$ this normal lies in a cone called the *normal cone*. The final piece of geometry comes from realising that, as the normal swings around from being one generator to another of this normal cone (as in Fig. 2.17), the small element of the corresponding manifold $\phi = 0$ itself envelops the *dual* of the cone. If you cannot swallow this geometrically, it is easy to see analytically. Taking P to be the origin without loss of generality, the envelope of

$$\sum_{i=1}^{m} x_i\xi_i = 0 \tag{2.54}$$

as ξ_i varies with $Q(\xi_1,\ldots,\xi_m) = 0$ is given by

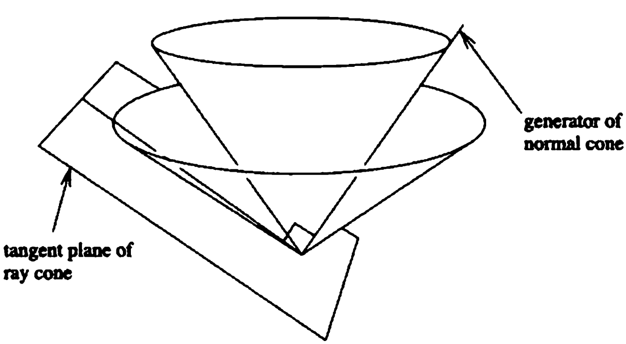


Fig. 2.17 Normal and ray cones.

$$\frac{\partial}{\partial \xi_i}\left(\sum_{j=1}^{m} x_j \xi_j\right) - \mu \frac{\partial Q}{\partial \xi_i} = 0 \quad \text{for } i = 1, \ldots, m, \tag{2.55}$$

where μ is a Lagrange multiplier. Hence we obtain

$$x_i = \mu \frac{\partial Q}{\partial \xi_i}, \quad Q(\xi_1, \ldots, \xi_m) = 0 \tag{2.56}$$

as the parametric representation of the envelope. It is clearly a homogeneous expression of degree n, this time in our base coordinates x_i, and thus also represents a cone, called the *ray cone*. To visualise the characteristic surface, all we have to think of is the ray cones dotted around in $\mathbb{R}^m$; a surface is characteristic if it is tangent to each ray cone at every point. For scalar equations, in which $n = 1$, the ray cones all collapse into lines, as in Fig. 1.2.

Once this picture has emerged, the concept of hyperbolicity takes on an entirely new flavour, because we no longer have a slope λ whose reality or otherwise could be used as a criterion. Instead, we are motivated to consider the possible geometry of the normal and ray cones and ask how degenerate they are. For example, if $Q(\xi_1, \ldots, \xi_m)$ is positive definite, the null vector is the only generator of the normal cone. We hope it is clear that the cones that have the greatest 'structure' are those for which there are n sheets all separated from each other; of course, some of these sheets might be flat (hyperplanes) or lines (one-dimensional).

The surprising result, which can only be verified by obtaining certain integral estimates on the bounds for $\mathbf{u}$, is that the Cauchy problem for the system (2.51) only turns out to be well posed in the sense defined after (2.24) when the normal and ray cones have the maximum structure they could have. To be precise, the Cauchy data must be prescribed on what is called a *space-like* hypersurface. Such hypersurfaces only exist if the ray cone has the maximal number of sheets that it could have and, in addition, all the sheets are 'nested' inside each other. Then the space-like hypersurfaces lie outside the 'outer' sheet of the ray cone; directions pointing inside this outer sheet are called *time-like* vectors.

We can illustrate this idea by combining two gasdynamics models, the unsteady one-dimensional model (2.3) and (2.4), and the steady two-dimensional model (2.5)–(2.7) and, as in the example on p. 42, linearising about a uniform velocity $(U, 0)$ to give

$$\begin{gathered}
\frac{\partial \bar{u}}{\partial t} + U\frac{\partial \bar{u}}{\partial x} = -\frac{a_0^2}{\rho_0}\frac{\partial \bar{\rho}}{\partial x}, \qquad \frac{\partial \bar{v}}{\partial t} + U\frac{\partial \bar{v}}{\partial x} = -\frac{a_0^2}{\rho_0}\frac{\partial \bar{\rho}}{\partial y}, \\
\frac{\partial \bar{\rho}}{\partial t} + U\frac{\partial \bar{\rho}}{\partial x} + \rho_0 a_0^2\left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y}\right) = 0.
\end{gathered} \tag{2.57}$$

When $U > 0$, the cones in (x, y, t) space are as in Figs 2.18 and 2.19; the (x, y) plane is always space-like but, when $U > a_0$, the time axis is no longer time-like![27]

[27]The same phenomenon occurs in the theory of relativity, when a particle travels faster than the speed of light in the material through which it is moving; it emits what is called Čerenkov radiation.

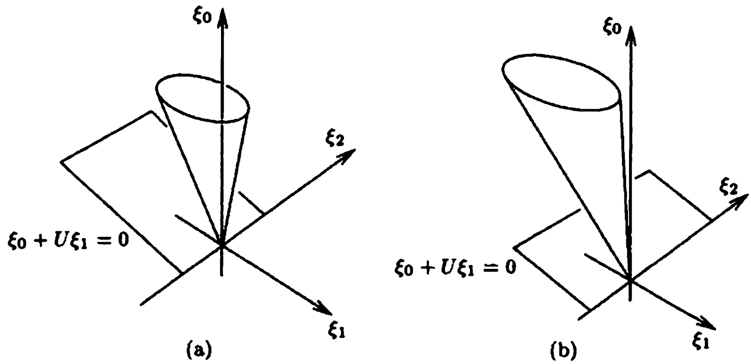


Fig. 2.18 Normal cones for (2.57) when (a) $U < a_0$ and (b) $U > a_0$, with $\xi_0 = t$, $\xi_1 = x$ and $\xi_2 = y$, so that $(\xi_0 + U\xi_1)\left((\xi_0 + U\xi_1)^2 - a_0^2(\xi_1^2 + \xi_2^2)\right) = 0$.

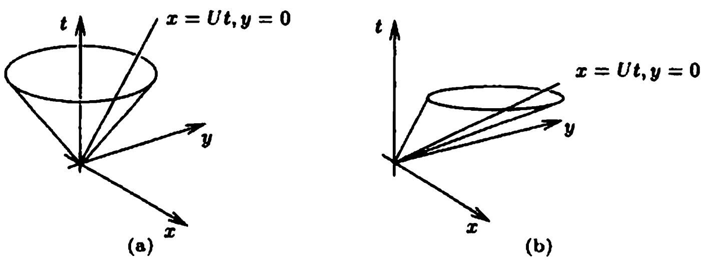


Fig. 2.19 Ray cones for (2.57) when (a) $U < a_0$ and (b) $U > a_0$, the union of $x = Ut$, $y = 0$ and $(x - Ut)^2 + y^2 = a_0^2 t^2$.

Another new twist comes when we consider the alternative definition of characteristics as surfaces across which $\partial \mathbf{u}/\partial x_i$ could have jump discontinuities. This still works fine, and leads back to (2.53), because all we need to say is

$$\sum_{i=1}^{m} \mathbf{A}_i \left[\frac{\partial \mathbf{u}}{\partial x_i}\right]_-^+ = \mathbf{0} \tag{2.58}$$

and

$$\sum_{i=1}^{m} \left[\frac{\partial \mathbf{u}}{\partial x_i}\right]_-^+ \delta x_i = \mathbf{0}$$

for all δx_i such that $\sum_{i=1}^{m} \xi_i \delta x_i = 0$. But this leaves open the question of the precise behaviour of such a discontinuous solution within the characteristic surface. We saw in Chapter 1 that if a discontinuity existed at any point of a characteristic then it was forced to remain non-zero at all points of that characteristic in the absence of any pathologies. To generalise this calculation to the situation here is tedious but, bearing in mind (2.58), we can see that the change of any jump $[\partial\mathbf{u}/\partial x_i]$ across the characteristic surface in a direction $\delta\mathbf{x}$ tangential to that surface is

$$\sum_{j=1}^{m}\left[\frac{\partial^2\mathbf{u}}{\partial x_i\partial x_j}\right]_-^+\delta x_j.$$

However, the partial differential equation gives

$$\sum_{j=1}^{m}\mathbf{A}_j\left[\frac{\partial^2\mathbf{u}}{\partial x_j\partial x_i}\right]_-^+ + \quad \text{terms linear in } \left[\frac{\partial\mathbf{u}}{\partial x_i}\right]_-^+ = \mathbf{0}.$$

Now we can only obtain information about $[\partial\mathbf{u}/\partial x_i]_-^+$ by suitably combining the second-order derivatives, and this clearly gives that the tangential derivative of the jump of some linear combination of $\partial\mathbf{u}/\partial x_i$ is only expressible in terms of that jump when we take the derivative along the generator of the ray cone tangent to the surface. This result generalises (1.34) and, with the jargon introduced there, the curve formed by these generators is called the *bicharacteristic*. Intuitively, we can think of bicharacteristics as the only curves in the characteristic manifold that can transmit information. We give a physical interpretation of this in Chapter 4 and describe how to construct the bicharacteristics in §8.2.6.

Rankine–Hugoniot conditions can be written down for weak solutions of higher-dimensional equations in conservation form, the simplest system being

$$\frac{\partial\mathbf{u}}{\partial x_1} + \tilde{\nabla}\cdot\mathbf{v}(\mathbf{u}) = \mathbf{0}, \tag{2.59}$$

where $\tilde{\nabla} = (\partial/\partial x_2, \ldots, \partial/\partial x_m)$ and $\mathbf{u}$ may be either a scalar or a vector, $\mathbf{v}$ being a vector or $(m-1)\times n$ matrix function of $\mathbf{u}$, respectively; in the latter case, $\tilde{\nabla}\cdot\mathbf{v}$ is interpreted as the column vector formed by taking the divergences of the rows with respect to $x_2, \ldots, x_m$. Then (2.45) generalises to

$$\left[\mathbf{v}\cdot\tilde{\nabla}\Phi\right]_-^+ + \frac{\partial\Phi}{\partial x_1}[\mathbf{u}]_-^+ = 0 \tag{2.60}$$

across a shock manifold defined by $\Phi(x_1, \ldots, x_m) = 0$.

We conclude with a surprisingly negative remark that can be made about *non-analytic* partial differential equations with three or more independent variables. Recall that, in the Cauchy–Riemann system, we have already displayed an equation with two independent variables for which there is no solution in the neighbourhood of a point at which non-analytic Cauchy data was prescribed. Now we can transfer this result into the statement that the inhomogeneous Cauchy–Riemann equations

$$\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = f'(x), \quad \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = 0, \tag{2.61}$$

with

$$u = v = 0 \quad \text{on } y = 0$$

have no analytic solution in a neighbourhood of $y = 0$ unless f is real and analytic. To see this, note that $u - f + \mathrm{i}v$ is an analytic function whose real part is $-f$ on $y = 0$ and whose imaginary part is zero on $y = 0$.

This is already a striking result, but it only ensures non-existence of solutions of (2.61) *locally* near $y = 0$. However, it is the starting point for H. Lewy's famous demonstration that, even though $f(x, y, t)$ may be C^∞ at every point in $\mathbb{R}^3$, the linear partial differential equation

$$\frac{\partial u}{\partial x} + \mathrm{i}\frac{\partial u}{\partial y} - 2\mathrm{i}(x + \mathrm{i}y)\frac{\partial u}{\partial t} = f(x, y, t)$$

has no solution at all unless f is in addition analytic. The details are too lengthy to reproduce here and we refer any reader who wishes to see the futility of trying to construct an all-embracing theory of partial differential equations to [23, 36].

Exercises

2.1. Show that, if $\gamma = 2$ and p/ρ^2 is constant, the gasdynamic equations (2.3) and (2.4) reduce to the shallow water equations (2.1) and (2.2), with ρ in (2.3) and (2.4) proportional to h in (2.1) and (2.2).

2.2. Show that when the substitution proposed before (2.8) is made in (2.5)–(2.7) and squares of barred quantities are neglected, we obtain

$$\frac{\partial}{\partial x}(\rho_0\bar{u} + U\bar{\rho}) + \frac{\partial}{\partial y}(\rho_0\bar{v}) = 0, \quad U\frac{\partial \bar{u}}{\partial x} = -\rho_0\frac{\partial \bar{p}}{\partial x}, \quad U\frac{\partial \bar{v}}{\partial x} = -\rho_0\frac{\partial \bar{p}}{\partial y}.$$

Show also that

$$\frac{p}{\rho^\gamma} = \frac{p_0}{\rho_0^\gamma}\left(1 + \left(\frac{\bar{p}}{p_0} - \gamma\frac{\bar{\rho}}{\rho_0}\right)\right)$$

in this approximation and hence deduce (2.8)–(2.10).

2.3. In a granular material, let the forces on the sides of a rectangle of area $\delta x \delta y$ be $(\sigma_x, \tau)\,\delta y$ on a side normal to the x axis and $(\tau, \sigma_y)\,\delta x$ on a side normal to the y axis, as in Fig. 2.20. Show that the forces on the diagonal are

$$\boldsymbol{\sigma}\,\delta s = (\sigma_x\,\delta y + \tau\,\delta x, \sigma_y\,\delta x + \tau\,\delta y)$$

and hence that

$$\boldsymbol{\sigma} = (\sigma_x \cos\theta + \tau\sin\theta, \sigma_y\sin\theta + \tau\cos\theta).$$

Now define

$$p = -\frac{\sigma_x + \sigma_y}{2}, \quad \tan 2\psi = \frac{2\tau}{\sigma_x - \sigma_y}.$$

Show that the components of $\boldsymbol{\sigma}$ along and perpendicular to $\mathbf{n}$ are

$$\sigma_n = -p + \tau\frac{\cos 2(\psi - \theta)}{\sin 2\psi}, \quad \tau_\theta = \tau\frac{\sin 2(\psi - \theta)}{\sin 2\psi},$$

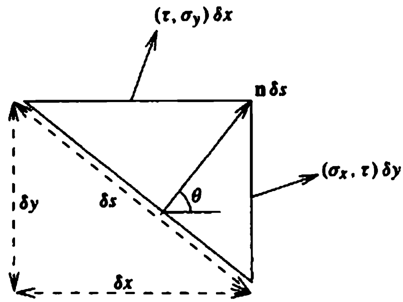


Fig. 2.20 Forces on an element of a granular medium.

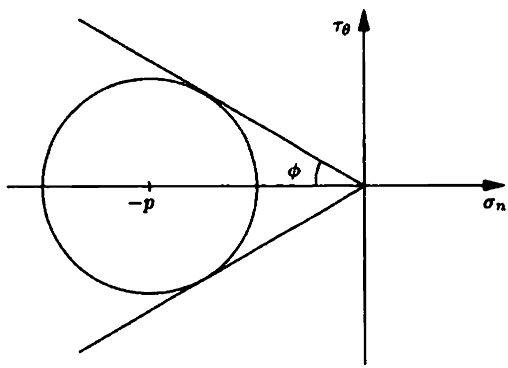


Fig. 2.21 The Mohr circle.

respectively. Deduce that

$$(\sigma_n + p)^2 + \tau_\theta^2 = \tau^2 \operatorname{cosec}^2 2\psi,$$

a relation which is known as the *Mohr circle* in the (σ_n, τ_θ) plane (see Fig. 2.21). Identify the region in this figure that corresponds to the friction constraint $|\tau_\theta/\sigma_n| \leqslant \tan\phi$, show that in limiting equilibrium

$$\tau = p \sin\phi \sin 2\psi, \quad \sigma_x = -p(1 - \sin\phi\cos 2\psi), \quad \sigma_y = -p(1 + \sin\phi\cos 2\psi),$$

and hence deduce (2.16) when $p > 0$.

2.4. (i) The discussion on p. 46 is equivalent to writing

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0$$

as a first-order system of dimension 2 by setting

$$\frac{\partial\phi}{\partial x} = u, \quad \frac{\partial\phi}{\partial y} = -v.$$

Can

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \phi = 0$$

be written as a system of dimension 2?

(ii) Show that, if $\mathbf{A}$ and $\mathbf{B}$ are constant matrices and

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

then

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{0}$$

can be written as a scalar second-order equation for u_1 or u_2. Would the same be true for a system with non-constant coefficients $\mathbf{A}$ and $\mathbf{B}$?

2.5. Consider the non-autonomous inhomogeneous Cauchy problem

$$\frac{\partial \mathbf{u}}{\partial x} = \mathbf{D}(\mathbf{u}, x, y)\frac{\partial \mathbf{u}}{\partial y} + \mathbf{d}(\mathbf{u}, x, y),$$

with

$$\mathbf{u} = \mathbf{u}_0 \quad \text{on } x = 0.$$

By writing

$$\frac{\partial \eta}{\partial x} = 0 \qquad \text{with} \qquad \eta = y \quad \text{on } x = 0$$

and

$$\frac{\partial \xi}{\partial x} = \frac{\partial \eta}{\partial y} \qquad \text{with} \qquad \xi = 0 \quad \text{on } x = 0,$$

and setting

$$\mathbf{v} = \begin{pmatrix} \mathbf{u} \\ \xi \\ \eta \end{pmatrix},$$

show that $\mathbf{v}$ satisfies the autonomous homogeneous Cauchy problem

$$\frac{\partial \mathbf{v}}{\partial x} = \mathcal{D}(\mathbf{v})\frac{\partial \mathbf{v}}{\partial y},$$

with

$$\mathbf{v} = \mathbf{v}_0 \quad \text{on } x = 0,$$

where

$$\mathcal{D} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{d} \\ \mathbf{0}^{\mathsf{T}} & 0 & 1 \\ \mathbf{0}^{\mathsf{T}} & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_0 = \begin{pmatrix} \mathbf{u}_0 \\ 0 \\ y \end{pmatrix}.$$

2.6. (i) Show that the solution of

$$\frac{\partial u}{\partial x} = u\frac{\partial u}{\partial y},$$

with $u(0, y) = y$, is $u = y/(1 - x)$.

(ii) For this example, in the notation of §2.3, $d(u) = u$ and $u_0(y) = y$. What are R_a and R_d? Take $\rho_a = \rho_d = \rho > 0$ and show that $M_a = M_d = \rho$. Deduce that $D(U) = \rho^2/(\rho - U)$ and $U_0(y) = \rho y/(\rho - y)$, and hence that the solution of $\partial U/\partial x = D(U)\,\partial U/\partial y$ is given implicitly by

$$y = \frac{-\rho^2 x}{\rho - U} + \frac{\rho U}{\rho + U}.$$

Deduce that U blows up on the lower branch of the parabola

$$y = \frac{\rho}{2}\left(1 - x \pm 2\sqrt{x}\right).$$

The distance from the nearest point on this curve to the origin controls the radius of convergence of (2.35); how does this distance depend on ρ?

2.7. Suppose that

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{c},$$

and $\mathbf{A}$ is invertible. Suppose also that the row vector $\boldsymbol{\ell}^{\mathsf{T}}$ is such that

$$\boldsymbol{\ell}^{\mathsf{T}}(\dot{y}\mathbf{I} - \dot{x}\mathbf{A}^{-1}\mathbf{B}) = \mathbf{0}^{\mathsf{T}}, \tag{2.62}$$

where $(\dot{x}, \dot{y})$ is such that $(x(t), y(t))$ is a characteristic. Show that $\boldsymbol{\ell}^{\mathsf{T}}\dot{\mathbf{u}} = \boldsymbol{\ell}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{c}\dot{x}$, which is (2.27).

2.8. Suppose that a hyperbolic system can be written in terms of the Riemann invariants $r_{1,2}$ as

$$\left(\frac{\partial}{\partial x} + \lambda_1(r_1, r_2)\frac{\partial}{\partial y}\right) r_1 = 0, \qquad \left(\frac{\partial}{\partial x} + \lambda_2(r_1, r_2)\frac{\partial}{\partial y}\right) r_2 = 0.$$

Show that x and y satisfy the *linear* system

$$\frac{\partial y}{\partial r_2} = \lambda_1(r_1, r_2)\frac{\partial x}{\partial r_2}, \qquad \frac{\partial y}{\partial r_1} = \lambda_2(r_1, r_2)\frac{\partial x}{\partial r_1}.$$

2.9. Write the system

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = 0,$$

$$\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} + 2\frac{\partial v}{\partial y} + \frac{\partial w}{\partial y} = 0,$$

$$\frac{\partial w}{\partial x} - \frac{\partial u}{\partial y} + 2\frac{\partial v}{\partial y} = 0,$$

in the form

$$\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{0},$$

and show that it is hyperbolic with $\mathrm{d}y/\mathrm{d}x = \lambda = \pm 1, 3$ on the characteristics.

Find the left eigenvalues of $\mathbf{B}$ and obtain the Riemann invariants $3u - v - w$, $u - v + w$ and $u + 3v + w$.

*2.10. Rearrange (2.3) and (2.4) into the form

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0, \quad \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0, \quad \frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + \gamma p\frac{\partial u}{\partial x} = 0,$$

which could also be written as $\mathbf{A}\,\partial\mathbf{u}/\partial x + \mathbf{B}\,\partial\mathbf{u}/\partial y = \mathbf{0}$, and show that the system is hyperbolic and that the characteristic speed $\mathrm{d}x/\mathrm{d}t = \lambda$ satisfies $\lambda = u$ or $\lambda = u \pm a$, where $a^2 = \gamma p/\rho$ (a is the sound speed). By multiplying by suitable row vectors $\boldsymbol{\ell}^{\mathsf{T}}$ satisfying $\boldsymbol{\ell}^{\mathsf{T}}(\dot{y}\mathbf{I} - \dot{x}\mathbf{A}^{-1}\mathbf{B}) = \mathbf{0}^{\mathsf{T}}$, obtain the relations

$$\gamma p\dot{\rho} - \rho\dot{p} = 0, \quad \pm\gamma p\dot{u} - a\dot{p} = 0$$

on the characteristics, respectively, and deduce that there is one Riemann invariant, namely p/ρ^{γ} on $\mathrm{d}x/\mathrm{d}t = u$.

Assuming that p/ρ^{γ} is the same constant everywhere, verify that the quantities $u \pm 2a/(\gamma - 1)$ are constant on $\mathrm{d}x/\mathrm{d}t = u \pm a$, respectively.

*2.11. Repeat Exercise 2.10 for the system (2.5)–(2.7) by rearranging the equations into the form $\mathbf{A}\,\partial\mathbf{u}/\partial x + \mathbf{B}\,\partial\mathbf{u}/\partial y = \mathbf{0}$, with $\mathbf{u} = (\rho, u, v, p)^{\mathsf{T}}$ and

$$\mathbf{A} = \begin{pmatrix} u & \rho & 0 & 0 \\ 0 & u & 0 & 1/\rho \\ 0 & 0 & u & 0 \\ 0 & \gamma p & 0 & u \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} v & 0 & \rho & 0 \\ 0 & v & 0 & 0 \\ 0 & 0 & v & 1/\rho \\ 0 & 0 & \gamma p & v \end{pmatrix}.$$

Deduce that the characteristic directions are given by $\mathrm{d}y/\mathrm{d}x = \lambda$, where $\lambda = v/u$ (twice) or

$$\lambda = -uv \pm \frac{a\sqrt{u^2 + v^2 - a^2}}{a^2 - u^2}, \quad a^2 = \frac{\gamma p}{\rho}.$$

Show that these results agree with the formulæ for λ on p. 53. Also obtain the relations

$$\gamma p\dot{\rho} - \rho\dot{p} = 0 \quad \text{and} \quad u\dot{u} + v\dot{v} + \frac{\dot{p}}{\rho} = 0$$

on $\lambda = v/u$, and

$$\gamma p(\lambda\dot{u} - \dot{v}) + (v - \lambda u)\dot{p} = 0$$

on the other characteristics, and find the Riemann invariants, if any exist.

2.12. Write the system of equations

$$u\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = v^{1/2}u,$$

$$v\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = u^{1/2}v$$

in the form

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{c}. \tag{2.63}$$

Assuming that u and v are strictly positive, show the characteristic directions are given by $\lambda = \dot{y}/\dot{x} = \pm(uv)^{-1/2}$. Show also that the corresponding left eigenvectors of Exercise 2.7 are $(v^{1/2}, \pm u^{1/2})$ and that, when $\lambda = 1/\sqrt{uv}$, (2.27) becomes

$$v^{1/2}\dot{u} + u^{1/2}\dot{v} = 2(uv)^{1/2}\dot{x}.$$

Show that the condition

$$\frac{\partial}{\partial v}\left(\frac{P}{R}\right) = \frac{\partial}{\partial u}\left(\frac{Q}{R}\right)$$

for (2.27) is satisfied, and finally that

$$u^{1/2} + v^{1/2} = x + \text{constant}$$

on this characteristic.

2.13. The system (2.1) and (2.2), with h proportional to v^2, can be written as

$$\frac{\partial u}{\partial x} + u\frac{\partial u}{\partial y} + 2v\frac{\partial v}{\partial y} = 0,$$

$$2\frac{\partial v}{\partial x} + 2u\frac{\partial v}{\partial y} + v\frac{\partial u}{\partial y} = 0.$$

Show that this system is hyperbolic with characteristic directions $\lambda = \mathrm{d}y/\mathrm{d}x = u \pm v$, and that $u \pm 2v$ are the corresponding Riemann invariants.
If $u(0, y) = 0$ and $v(0, y)$ is continuous and takes the value 1 in $y < 0$, show that $u = 0$ and $v = 1$ in $x + y < 0$, $x > 0$, assuming that no shocks form. Show also that, for $x + y$ sufficiently small and positive,

$$\frac{\partial u}{\partial x} + \left(\frac{3u}{2} - 1\right)\frac{\partial u}{\partial y} = 0,$$

with $u = 0$ on $x + y = 0$. Verify that the equation is satisfied by

$$u = \frac{2(x+y)}{3x}.$$

What properties must $v(0, y)$ have for $y > 0$ for this solution to be relevant?

*2.14. To ease the calculations in Exercise 2.3, set $\phi = \pi/2$ and $p = -(\sigma_x + \sigma_y)/2$. Using the relations

$$\tau = p\sin 2\psi, \quad \tan 2\psi = \frac{2\tau}{\sigma_x - \sigma_y},$$

show that

$$\sigma_x = -p(1 - \cos 2\psi), \quad \sigma_y = -p(1 + \cos 2\psi),$$

and hence that

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{0},$$

where $\mathbf{u} = (p, \psi)^{\mathsf{T}}$ and

$$\mathbf{A} = \begin{pmatrix} \cos 2\psi - 1 & -2p\sin 2\psi \\ \sin 2\psi & 2p\cos 2\psi \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \sin 2\psi & 2p\cos 2\psi \\ -(1+\cos 2\psi) & 2p\sin 2\psi \end{pmatrix}.$$

Deduce that the system has a double characteristic

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lambda = -\cot\psi,$$

on which ψ is a constant. (If $\phi < \pi/2$, it can be shown that the characteristics split into

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\cot\left(\psi \pm \left(\frac{\pi}{4} - \frac{\phi}{2}\right)\right)$$

and the Riemann invariants are $\psi \pm \frac{1}{2}\cot\phi \log p$.)

2.15. Suppose that A and τ are related to x and t by

$$A = A(x,t) \quad \text{and} \quad \tau = \int_{t_0}^{t} T(t')\,\mathrm{d}t'.$$

Using the chain rule

$$\frac{\partial}{\partial x} \to \frac{\partial A}{\partial x}\frac{\partial}{\partial A}, \quad \frac{\partial}{\partial t} \to T\frac{\partial}{\partial \tau} + \frac{\partial A}{\partial t}\frac{\partial}{\partial A},$$

show that

$$\frac{\partial A}{\partial x} = 1\Big/\frac{\partial x}{\partial A} \quad \text{and} \quad \frac{\partial A}{\partial t} = -T\frac{\partial x}{\partial \tau}\Big/\frac{\partial x}{\partial A}.$$

Hence eliminate u between (2.11) and (2.12) to show that

$$\frac{\partial^2 x}{\partial A^2} - c\frac{\partial^2 x}{\partial A \partial \tau} = -\frac{1}{A}\frac{\partial x}{\partial A}.$$

2.16. The system

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{1}{h^2}\frac{\partial h}{\partial x}, \qquad \frac{\partial (h^2)}{\partial t} + \frac{\partial (h^2 v)}{\partial x} = 0$$

is an approximate model for a slender axisymmetric inviscid fluid jet of radius $h(x,t)$ and axial velocity $v(x,t)$, aligned with the x axis. Show that, if x and t are regarded as functions of v and h, the system becomes

$$\frac{\partial x}{\partial h} - v\frac{\partial t}{\partial h} - \frac{1}{h^2}\frac{\partial t}{\partial v} = 0, \qquad \frac{h}{2}\frac{\partial t}{\partial h} - v\frac{\partial t}{\partial v} + \frac{\partial x}{\partial v} = 0.$$

Show further that eliminating x leads to

$$\frac{\partial}{\partial h}\left(h^3\frac{\partial t}{\partial h}\right) + 2\frac{\partial^2 t}{\partial v^2} = 0.$$

2.17. The system (2.3) and (2.4) is derived from the physical conservation laws

$$\frac{\partial P}{\partial t} + \frac{\partial Q}{\partial x} = 0,$$

where

(i) $P = \rho$, $Q = \rho u$;

(ii) $P = \rho u$, $Q = p + \rho u^2$;

(iii) $P = p/(\gamma - 1) + \frac{1}{2}\rho u^2$, $Q = u\left(p/(\gamma - 1) + \frac{1}{2}\rho u^2\right) + pu$.

Show that the Rankine–Hugoniot relation for a shock moving with speed V is $[Q]_-^+ = V[P]_-^+$, and hence from (i) that $[\rho u]_-^+ = V[\rho]_-^+$. Use this with (ii) and (iii) to deduce the remaining conditions in (2.49).

2.18. Look for a travelling wave solution $u(x, y) = U(z)$, with $z = (y - Vx)/\epsilon$, to the equation

$$\frac{\partial u}{\partial x} + \frac{\partial}{\partial y}\left(\frac{1}{n}u^n\right) = \epsilon\frac{\partial^2 u}{\partial y^2} \quad \text{for } -\infty < z < \infty.$$

Show that

$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{1}{n}U^n - VU + \text{constant},$$

and deduce that $V = \left[\frac{1}{n}U^n\right]_{-\infty}^{\infty} / [U]_{-\infty}^{\infty}$. Show also that, when $n = 2$, U can only tend to $U(\pm\infty)$ as $z \to \pm\infty$ if $\mathrm{d}U/\mathrm{d}z < 0$.

3

Introduction to second-order scalar equations

3.1 Preamble

In Chapter 2 we tried to describe, in the simplest possible terms, the only reasonably general framework that exists concerning the solutions of the Cauchy problem for arbitrary partial differential equations. Many pages could now be spent in explaining this framework in more detail, but we will instead devote the next three chapters to a more specialised analysis of three commonly occurring classes of *second-order scalar* equations. This will enable us not only to illustrate as simply as possible the subtleties that can arise when 'non-hyperbolic' equations are being solved, but also to look at the role played by the ideas proposed in Chapter 2 when they are viewed in a more concrete setting.

When we start to consider specific second-order equations in Chapters 4–6, we will find that we rapidly encounter formidable technical difficulties. Hence this chapter is a preface that emphasises the central ideas that will need to be kept in mind in the next three chapters.

Before we start, let us make one elementary observation to illustrate the care that may have to be taken with what appear to be harmless enough second-order scalar equations. It comes from a trivial piece of Fourier analysis and shows that the qualitative behaviour of the solutions of second-order partial differential equations with Cauchy data is far more diverse than in the case for initial value problems for second-order ordinary differential equations. In the latter case, the Cauchy–Picard theorem guarantees existence, uniqueness and continuous dependence on the data, assuming only appropriate Lipschitz continuity. However, we shall see that partial differential equations which are on the face of it much smoother can in fact lose the property of continuous dependence on the Cauchy data. This fact, which provides one of the basic motivations for trying to classify these partial differential equations, is illustrated by considering the solutions of the two problems

(i)
$$\frac{\partial u}{\partial y} = \frac{\partial^2 u}{\partial x^2}, \tag{3.1}$$

(ii)
$$\frac{\partial u}{\partial y} = -\frac{\partial^2 u}{\partial x^2} \tag{3.2}$$

in $y > 0$, with data

$$u(x,0) = \sum_{n=0}^{\infty} a_n \cos nx \quad \text{for } -\infty < x < \infty. \tag{3.3}$$

We assume that a_n are such that $u(x,0)$ has continuous second derivatives, which clearly demands that $a_n = o(n^{-2})$ as $n \to \infty$.

Separation of the variables gives the respective solutions as

(i)
$$u = \sum_{n=0}^{\infty} a_n e^{-n^2 y} \cos nx, \tag{3.4}$$

(ii)
$$u = \sum_{n=0}^{\infty} a_n e^{n^2 y} \cos nx. \tag{3.5}$$

The series (3.4) has better convergence properties than (3.3) for $y > 0$, but the series (3.5) does not even converge unless a_n decays very rapidly indeed as $n \to \infty$. For example, even if $a_n = e^{-n^2}$, (3.5) blows up as $y \to 1$ (if $a_n = e^{-\epsilon n^2}$, $\epsilon > 0$, blow-up is at $y = \epsilon$, however small that may be, and if $a_n = O(e^{-\epsilon n})$, as is generic for periodic analytic functions, there is no solution at all).

Similarly, if we consider the problems

(iii)
$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial x^2}, \tag{3.6}$$

(iv)
$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 u}{\partial x^2} \tag{3.7}$$

for $y > 0$, with, for example, data

$$u(x,0) = \sum_{n=0}^{\infty} a_n \cos nx, \quad \frac{\partial u}{\partial y}(x,0) = 0, \tag{3.8}$$

the Fourier series solutions are

(iii)
$$\sum_{n=0}^{\infty} a_n \cos ny \cos nx, \tag{3.9}$$

(iv)
$$\sum_{n=0}^{\infty} a_n \cosh ny \cos nx, \tag{3.10}$$

respectively, and a similar contrast can be drawn to that between (3.4) and (3.5).

The switch from (3.6) to (3.7) can easily occur in practice, as has been anticipated on p. 42. Suppose that we have an 'irrotational' solution of the system (2.8)–(2.10) describing linearised steady gas dynamics, so that $(\bar{u}, \bar{v}) = \nabla\phi$. Eliminating $\bar{\rho}$ gives

$$\left(1 - \frac{U^2}{a_0^2}\right) \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0; \tag{3.11}$$

depending whether the Mach number U/a_0 is greater or less than unity, we are effectively led to (3.6) or (3.7), respectively. In fact, the derivation of (3.11) breaks down when U/a_0 is too close to unity, and in some 'transonic' flows the appropriate model is

$$\frac{\partial \phi}{\partial x}\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0. \tag{3.12}$$

Now, as in (2.41), we can make the hodograph transformation

$$x = X(U, V), \quad y = Y(U, V),$$

where

$$U = \frac{\partial \phi}{\partial x}, \quad V = \frac{\partial \phi}{\partial y},$$

to give

$$U\frac{\partial^2 Y}{\partial V^2} + \frac{\partial^2 Y}{\partial U^2} = 0. \tag{3.13}$$

This equation is called the *Tricomi equation*; it is more subtle than (3.6) or (3.7), and we will return to it later.

A final motivation for studying the contrast between (3.6) and (3.7) comes from our glass manufacture model (2.13)–(2.15). When we try to predict the centre surface $z = H(x, y)$, we have to consider a *lateral* force balance in Fig. 2.4, which can be shown to give

$$\sigma_x \frac{\partial^2 H}{\partial x^2} + 2\tau \frac{\partial^2 H}{\partial x \partial y} + \sigma_y \frac{\partial^2 H}{\partial y^2} = 0, \tag{3.14}$$

where σ_x, σ_y and τ are given in terms of the stretching velocity (u, v) by (2.14). To be as unambitious as possible, we just consider the simplest velocity distribution, which is linear in position with $u = \alpha x$ and $v = \beta y$; we obtain

$$(2\alpha + \beta)\frac{\partial^2 H}{\partial x^2} + (\alpha + 2\beta)\frac{\partial^2 H}{\partial y^2} = 0. \tag{3.15}$$

Hence, depending on the signs of the coefficients, we again have a situation in which we see the fundamental importance of recognising the difference between (3.6) and (3.7).

With this preamble in mind, let us now consider the Cauchy problem for scalar second-order equations. It is much easier to begin with the semilinear case with just two independent variables and we defer consideration of genuinely quasilinear equations to the end of the chapter.

3.2 The Cauchy problem for semilinear equations

We consider second-order scalar problems with linear principal part in the form

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} = f\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right), \tag{3.16}$$

where a, b and c are functions of x and y. We can only use the transformation $u_1 = \partial u/\partial x$, $u_2 = \partial u/\partial y$ to reduce this to a two-dimensional system, as considered in Chapter 2, when the right-hand side of this equation does not depend on u (otherwise the resulting system is three-by-three). If this can be done, appropriate Cauchy boundary data would be for u_1 and u_2 to be given on some curve Γ in the (x, y) plane. However, even when f depends on u, it is reasonable to consider Cauchy boundary data for (3.16) in the form

$$x = x_0(s), \quad y = y_0(s), \quad \frac{\partial u}{\partial x} = p_0(s), \quad \frac{\partial u}{\partial y} = q_0(s) \quad \text{for } s_1 \leqslant s \leqslant s_2. \tag{3.17}$$

Now, by integrating along Γ, this implies that $u = u_0(s)$ is prescribed up to a constant, and, if we know this constant, an equivalent[28] set of data is

$$u = u_0(s), \quad \frac{\partial u}{\partial n} = v_0(s) \quad \text{for } s_1 \leqslant s \leqslant s_2, \tag{3.18}$$

and it is this that is traditionally called Cauchy data for (3.16). In geometrical terms the Cauchy data states that the solution surface $u = u(x, y)$ not only has to pass through a boundary curve whose projection is Γ but also has to have a prescribed tangent plane on that curve, so that a 'boundary strip' is given for this surface.

A necessary but not always sufficient condition for the existence of a solution of (3.16) is that the boundary data (3.18) defines the second derivatives $\partial^2 u/\partial x^2$, $\partial^2 u/\partial x \partial y$ and $\partial^2 u/\partial y^2$ uniquely on the boundary curve.[29] In addition to

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} = f,$$

differentiation along Γ shows that these derivatives must satisfy

$$p_0' = x_0'\frac{\partial^2 u}{\partial x^2} + y_0'\frac{\partial^2 u}{\partial x \partial y},$$

$$q_0' = x_0'\frac{\partial^2 u}{\partial x \partial y} + y_0'\frac{\partial^2 u}{\partial y^2}.$$

Thus the condition is

$$\begin{vmatrix} a & 2b & c \\ x_0' & y_0' & 0 \\ 0 & x_0' & y_0' \end{vmatrix} \neq 0,$$

which reduces to

$$ay_0'^2 - 2bx_0'y_0' + cx_0'^2 \neq 0. \tag{3.19}$$

This is the generalisation to second-order equations of (1.5) and it is equivalent to (2.23) in the case that (3.16) can be identified with a two-by-two system.

[28]This is because $\mathrm{d}u_0/\mathrm{d}s = p_0\,\mathrm{d}x_0/\mathrm{d}s + q_0\,\mathrm{d}y_0/\mathrm{d}s$ and $\partial u/\partial n = -q_0\,\mathrm{d}x_0/\mathrm{d}s + p_0\,\mathrm{d}y_0/\mathrm{d}s$ when we parametrise so that s is arc length.

[29]Of course, this says nothing about higher derivatives of u.

3.3 Characteristics

As in Chapters 1 and 2, we now define a characteristic of (3.16) to be a curve in the (x, y) plane along which Cauchy data does not uniquely define the second derivatives. Hence, given $x(t)$, $y(t)$, $\partial u/\partial x\,(t)$ and $\partial u/\partial y\,(t)$ on a characteristic, we can use the argument just given to state that

$$a\dot{y}^2 - 2b\dot{x}\dot{y} + c\dot{x}^2 = 0. \tag{3.20}$$

This defines the two characteristic directions

$$\frac{dy}{dx} = \frac{b \pm \sqrt{b^2 - ac}}{a}$$

at a point, but they may not be real or distinct. Moreover, for existence (and non-uniqueness) of the second derivatives, there is a further relation[30] which holds on a characteristic and can be written as

$$\begin{vmatrix} a & f & c \\ \dot{x} & \dot{p} & 0 \\ 0 & \dot{q} & \dot{y} \end{vmatrix} = 0,$$

where we have used the standard notation $p = \partial u/\partial x$ and $q = \partial u/\partial y$. This reduces to

$$a\dot{y}\dot{p} + c\dot{x}\dot{q} = f\dot{x}\dot{y}, \tag{3.21}$$

but, thought of as an ordinary differential equation, this equation is only rarely integrable with respect to t. However, if we are lucky enough to be able to integrate it along a real characteristic, a linear combination of $\partial u/\partial x$ and $\partial u/\partial y$ is known along that characteristic; this is what we called a Riemann invariant on p. 44. Thus, if we know a Riemann invariant on a family of characteristics, our second-order partial differential equation is reduced to a first-order one.

Better still, if *two* real distinct families of characteristics exist and there is a Riemann invariant for each of them, then it is possible to compute $\partial u/\partial x$ and $\partial u/\partial y$, and hence obtain u algebraically, from the two relations provided by the two Riemann invariants, and this is a key motivation for forthcoming definitions. It is helpful to note that, when $f = 0$, (3.20) and (3.21) imply that on a characteristic

$$a\dot{p}^2 + 2b\dot{p}\dot{q} + c\dot{q}^2 = 0. \tag{3.22}$$

We remark that, as in (1.17), the alternative definition of a characteristic as a curve across which there can be discontinuities in the *second* derivatives of u leads at once to (3.20) but not to (3.21).

When we now come to decide what role the characteristics play in the solution of a boundary value problem for (3.16), we note that, of the model problems discussed in §3.1, equations (3.6) and (3.7) with the data (3.8) are clearly Cauchy problems whose solutions have very different properties from each other, and that

[30] This is another consequence of the Fredholm Alternative.

(3.6) has real characteristics while (3.7) does not. We also note that (3.1) and (3.2) have the attributes of a Cauchy problem in that there are as many pieces of data as there are y derivatives in the partial differential equations, and that they have coincident real characteristics y = constant. A picture of how we should classify equations such as (3.16) is now emerging, and corroboration is provided by noting what happens to (3.16) if we change coordinates with the dual aim of identifying invariant properties of the left-hand side of the equation and, if possible, writing it in a simpler 'canonical form'. We therefore consider a one-to-one transformation $\xi = \xi(x, y)$, $\eta = \eta(x, y)$, for which the Jacobian $|\partial(\xi, \eta)/\partial(x, y)|$ is bounded and non-zero. From the chain rule we find that

$$\alpha\frac{\partial^2 u}{\partial\xi^2} + 2\beta\frac{\partial^2 u}{\partial\xi\partial\eta} + \gamma\frac{\partial^2 u}{\partial\eta^2} = g\left(\xi, \eta, u, \frac{\partial u}{\partial\xi}, \frac{\partial u}{\partial\eta}\right), \tag{3.23}$$

for some function g, where

$$\alpha = a\left(\frac{\partial\xi}{\partial x}\right)^2 + 2b\frac{\partial\xi}{\partial x}\frac{\partial\xi}{\partial y} + c\left(\frac{\partial\xi}{\partial y}\right)^2, \tag{3.24}$$

$$\beta = a\frac{\partial\xi}{\partial x}\frac{\partial\eta}{\partial x} + b\left(\frac{\partial\xi}{\partial x}\frac{\partial\eta}{\partial y} + \frac{\partial\xi}{\partial y}\frac{\partial\eta}{\partial x}\right) + c\frac{\partial\xi}{\partial y}\frac{\partial\eta}{\partial y}, \tag{3.25}$$

$$\gamma = a\left(\frac{\partial\eta}{\partial x}\right)^2 + 2b\frac{\partial\eta}{\partial x}\frac{\partial\eta}{\partial y} + c\left(\frac{\partial\eta}{\partial y}\right)^2. \tag{3.26}$$

Hence, in the new coordinates the discriminant of the characteristic equation (3.20) becomes

$$\beta^2 - \alpha\gamma = (b^2 - ac)\left|\frac{\partial(\xi, \eta)}{\partial(x, y)}\right|^2,$$

where $b^2 - ac$ is the discriminant in the old coordinates. Consequently, the reality or otherwise of characteristics of (3.16) is invariant under our transformation.

This is the final piece of evidence that motivates us to classify an equation of the form (3.16) as follows.

- It is *hyperbolic* if $b^2 > ac$. Then two real distinct characteristics exist and, if (3.9) is a guide, the solution to the Cauchy problem is 'well behaved'. This case will be considered in more detail in Chapter 4.
- It is *elliptic* if $b^2 < ac$. Then there are no real characteristics and, if (3.10) is a guide, the solution to the Cauchy problem is unpredictable to say the least. This case will be considered in more detail in Chapter 5.
- It is *parabolic* if $b^2 = ac$. Then there is one 'double' real characteristic and, if (3.4) and (3.5) are a guide, the solution of what we have called the Cauchy problem may or may not be well behaved. This is clearly the most delicate of the three situations, but it is also the commonest in practice, at least in models of industrial interest. This explains why Chapter 6, in which we consider parabolic problems in more detail, is so lengthy.

This classification immediately suggests that we should be able to find coordinate systems especially suitable for instant recognition of the type of any equation of the form (3.16), and we will indeed find such *canonical forms* shortly. However, let us first digress to note that the practical examples mentioned in §3.1 suggest some models that have the unfortunate property of being of *mixed type*, that is they may be hyperbolic in some region in the (x, y) plane and elliptic in others. In fact, it is a trivial exercise to show that the characteristics of the Tricomi equation (3.13) are as shown in Fig. 3.1 (see Exercise 3.5(b)); it is only hyperbolic in $U < 0$, which can be shown to be the region of supersonic flow, and this region is bounded by the *sonic line* $U = 0$. Equally, if we could justify allowing α and β to be functions of position in (3.15), the hyperbolicity of that equation would depend on how great the local extension and compression were. This sends us three clear messages: first, that we expect the physical implications of any mathematical solution to change dramatically as we cross any sonic line from elliptic to hyperbolic; second, that any boundary data that we propose in the hope of writing down a well-posed problem may well have to change abruptly wherever the boundary intersects a sonic line; third, that any numerical discretisation that we might use should reflect the change in behaviour. Moreover, there is another important aspect of problems of mixed type that we have rather glossed over. In the Tricomi model (3.13) and the glass sheet model (3.15), the position of the sonic line is known in advance. For the Tricomi equation, this was the result of the hodograph transformation and, for the glass model, the output of some uncoupled problem for the fluid velocity. In practice, however, the coefficients on the left-hand side of (3.16) usually involve the dependent variable in some way, and the system is 'clever' enough to switch type from, say, hyperbolic to elliptic, in such a way as not to exhibit any pathological behaviour such as (3.10). However, this discussion leads us away from the semilinear cases we are considering, for which we have still not addressed the question of canonical forms.

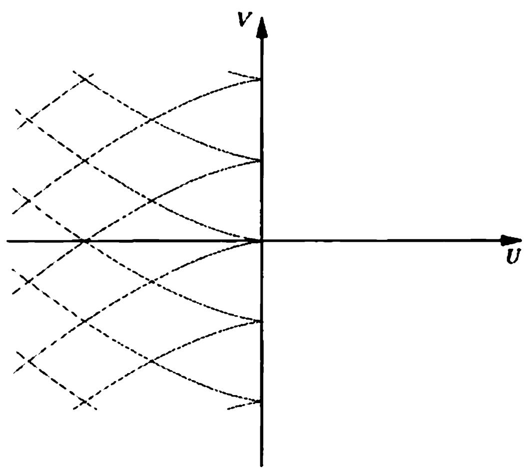


Fig. 3.1 Characteristics for the Tricomi equation (3.13).

3.4 Canonical forms for semilinear equations

3.4.1 Hyperbolic equations

When the equation

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} = f\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right)$$

is hyperbolic, two real characteristic directions exist at each point and define two distinct families of characteristics. Now, assuming Lipschitz continuity conditions on a, b and c, the characteristic condition (3.20), $a\dot{y}^2 - 2b\dot{x}\dot{y} + c\dot{x}^2 = 0$, is a pair of ordinary differential equations for which integrals certainly exist, although it may not be possible to write them down explicitly. Thus two *characteristic variables* may be defined, constant values of one variable giving one characteristic family, and constant values of the second variable giving the other. Such a definition is not unique, since any well-behaved function of a characteristic variable thus defined would also be a characteristic variable, and hence, in practice, such variables are often chosen aesthetically. All that is necessary is that we choose ξ and η in (3.23) so that $\xi(x, y) =$ constant and $\eta(x, y) =$ constant are in fact the integral curves of the characteristic equations.[31] Now, on $\xi(x, y) =$ constant, $(\partial\xi/\partial x)\dot{x} + (\partial\xi/\partial y)\dot{y} = 0$, with a similar relation on $\eta(x, y) =$ constant. Substituting into (3.20), we find that $(\partial\xi/\partial x) / (\partial\xi/\partial y)$ and $(\partial\eta/\partial x) / (\partial\eta/\partial y)$ are the two (real and distinct) roots of

$$a\lambda^2 + 2b\lambda + c = 0. \tag{3.27}$$

Hence, in (3.24)–(3.26), $\alpha = \gamma = 0$. Also

$$2\beta = a\left(\frac{\partial\xi}{\partial x} + \frac{\partial\eta}{\partial x}\right)^2 + 2b\left(\frac{\partial\xi}{\partial x} + \frac{\partial\eta}{\partial x}\right)\left(\frac{\partial\xi}{\partial y} + \frac{\partial\eta}{\partial y}\right) + c\left(\frac{\partial\xi}{\partial y} + \frac{\partial\eta}{\partial y}\right)^2,$$

and this quantity is non-zero since (3.27) only has two roots for λ. Thus a hyperbolic equation has the *canonical form*

$$\frac{\partial^2 u}{\partial\xi\partial\eta} = G\left(\xi, \eta, u, \frac{\partial u}{\partial\xi}, \frac{\partial u}{\partial\eta}\right). \tag{3.28}$$

Different choices of characteristic variables merely lead to different right-hand sides in this equation.

Example 3.1 (*The wave equation*) A very well-studied example leading to (3.28) is

$$a_0^2\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f(x, y), \quad a_0 = \text{constant}, \tag{3.29}$$

which, when $f = 0$, is referred to as the *wave equation*. The characteristics are given by $a_0^2\dot{y}^2 = \dot{x}^2$ and the problem is hyperbolic; one particular choice of characteristic

[31] This distinction is related to our discussion of the representation of characteristics in §§1.8 and 2.6.

variables is $\xi = x - a_0 y$ and $\eta = x + a_0 y$. It is easily verified that $\alpha = \gamma = 0$ and $\beta = 2a_0^2$, so that a canonical form is, say,

$$\frac{\partial^2 u}{\partial\xi\partial\eta} = \frac{1}{4a_0^2} f\left(\frac{\xi+\eta}{2}, \frac{\eta-\xi}{2a_0}\right) = g(\xi,\eta). \tag{3.30}$$

This example reveals one of the benefits that can occasionally accrue from writing the equation in canonical form; it actually enables us to perform an explicit integration to give the general solution of (3.30) as the *d'Alembert representation*

$$u = \iint g(\xi,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta + h_1(\xi) + h_2(\eta), \tag{3.31}$$

where h_1 and h_2 are arbitrary functions to be determined by the boundary conditions. Alternatively, we could have derived this formula by integrating (3.29) along the characteristics to give the Riemann invariants

$$\pm a_0 \frac{\partial u}{\partial x} - \frac{\partial u}{\partial y} = \int f\,\mathrm{d}y \quad \text{on } x \mp a_0 y = \text{constant}. \tag{3.32}$$

Example 3.2 Things are not usually so easy, as can be seen by considering

$$y^2\frac{\partial^2 u}{\partial x^2} - x^2\frac{\partial^2 u}{\partial y^2} = 0. \tag{3.33}$$

The characteristics are given by $y\dot{y} = \pm x\dot{x}$, and the equation is hyperbolic except on the axes where $x = 0$ or $y = 0$, which we exclude from the domain of interest. Characteristic variables are given by $\xi = y^2 - x^2$ and $\eta = x^2 + y^2$, and the two families of characteristics are shown in Fig. 3.2; note that they touch on the axes.

A tedious calculation leads to

$$\frac{\partial^2 u}{\partial\xi\partial\eta} = \frac{1}{2(\eta^2-\xi^2)}\left(\xi\frac{\partial u}{\partial\eta} - \eta\frac{\partial u}{\partial\xi}\right), \tag{3.34}$$

which is in canonical form, but there are no Riemann invariants, and neither can (3.34) be integrated explicitly.

3.4.2 Elliptic equations

In the elliptic case, real characteristic variables no longer exist and the integrals of (3.20) are complex conjugate pairs of functions $y(x)$. We choose ξ and η to be the real and imaginary parts of these complex integrals so that the 'complex' characteristics are $\xi \pm \mathrm{i}\eta = $ constant, and $(\partial\xi/\partial x \pm \mathrm{i}\partial\eta/\partial x)\dot{x} + (\partial\xi/\partial y \pm \mathrm{i}\partial\eta/\partial y)\dot{y} = 0$. Substituting into (3.23) and taking real and imaginary parts we obtain $\alpha = \gamma \neq 0$ and $\beta = 0$. The canonical form is therefore

$$\frac{\partial^2 u}{\partial\xi^2} + \frac{\partial^2 u}{\partial\eta^2} = G\left(\xi,\eta,u,\frac{\partial u}{\partial\xi},\frac{\partial u}{\partial\eta}\right). \tag{3.35}$$

When $G = 0$, this equation is called *Laplace's equation*. The Cauchy–Riemann system on p. 46 and the example (3.7) and (3.8) both show that Cauchy boundary

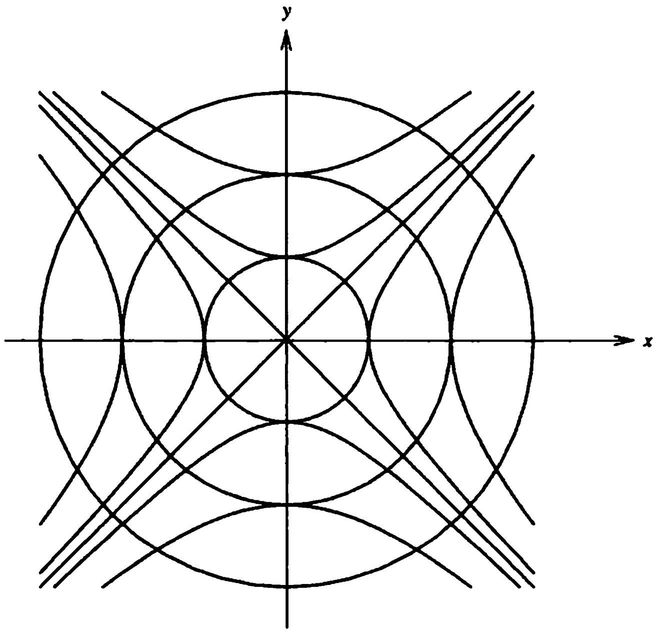


Fig. 3.2 Characteristics for equation (3.33).

data is inappropriate if (3.35) is to be well posed with $G = 0$. In this case we are simply restating the dangers about analytic continuation that we mentioned in §2.3, but the statement is in fact true for all elliptic equations, semilinear or otherwise. Hence, our first priority in Chapter 5 will be to decide what data is appropriate for elliptic equations. Apart from the obvious statement that the imposition of Cauchy data is too strong, the only general observation that can be made about (3.35) is when $G = 0$. The analogue of (3.31) is

$$u = h_1(\xi + \mathrm{i}\eta) + h_2(\xi - \mathrm{i}\eta), \tag{3.36}$$

where h_i must be analytic functions of their arguments for differentiation to be justified. Since we are implicitly seeking real solutions,[32] it is better to write (3.36) as

$$u = \Re\, h(\xi + \mathrm{i}\eta),$$

which is the 'general solution' of Laplace's equation, although we could equally write $u = \Re\, h(\xi - \mathrm{i}\eta)$. Note the contrast with (3.31), where the general solution

[32] In this context it is interesting to note that we can always write down an analytic function $h(\xi + \mathrm{i}\eta)$ when we know its real part $u(\xi, \eta)$; the formula is

$$h(\xi + \mathrm{i}\eta) = 2u\left(\frac{\xi + \mathrm{i}\eta}{2}, \frac{\xi - \mathrm{i}\eta}{2\mathrm{i}}\right) + \text{ constant}.$$

involved two twice-differentiable functions rather than one analytic one. Not surprisingly, the theory of functions of a complex variable is intimately related to the solution of Laplace's equation in $\mathbb{R}^2$ and Chapter 5 provides several illustrations of how this theory can be put to practical use.

3.4.3 Parabolic equations

In the parabolic case, the characteristic equation (3.20) has a double root and there is only one real characteristic variable. If we choose η to be this variable, so that $(\partial\eta/\partial x)\dot{x} + (\partial\eta/\partial y)\dot{y} = 0$, then, on substituting into (3.26), $\gamma = 0$. As always, there is some freedom of choice for ξ and, if $a \neq 0$, we may conveniently choose $\xi = x$ so that $\alpha = a$ (if $a = 0$ we choose $\xi = y$, since then $c \neq 0$). But $\gamma = (a\,\partial\eta/\partial x + b\,\partial\eta/\partial y)^2/\,a = \beta^2/a$ and hence $\beta = 0$. The canonical form is thus

$$\frac{\partial^2 u}{\partial \xi^2} = G\left(\xi, \eta, u, \frac{\partial u}{\partial \xi}, \frac{\partial u}{\partial \eta}\right). \tag{3.37}$$

The subtlety of this case is illustrated by contrasting the parabolic equations $\partial^2 u/\partial\xi^2 = 0$ and $\partial^2 u/\partial\xi^2 = \partial u/\partial\eta$. The former, although trivial, behaves like a hyperbolic equation because the general solution is $\xi f(\eta) + g(\eta)$, where f and g are arbitrary, and hence this solution is uniquely defined by Cauchy data on any non-characteristic, i.e. any line excluding $\eta =$ constant;[33] the latter, which is called the ***diffusion equation*** or ***heat equation***, is so complicated as to occupy much of Chapter 6.

We remark that, as often happens in applied mathematics, degenerate cases are best understood by taking judiciously chosen limits of non-degenerate ones. Here, a good idea of the kind of data that should be applied to parabolic equations can be gleaned from the limit of the hyperbolic equation

$$\frac{\partial^2 u}{\partial \xi^2} - \epsilon^2 \frac{\partial^2 u}{\partial \eta^2} = \frac{\partial u}{\partial \eta} \tag{3.38}$$

as $\epsilon \to 0$. This equation is an example of what will be studied in Chapter 4 as the ***telegraph equation***; its characteristics are $\eta \pm \epsilon\xi =$ constant, and we expect that well-posed data would be, for example, the prescription

$$u = f(\xi), \quad \frac{\partial u}{\partial \eta} = g(\xi)$$

on the non-characteristic $\eta = 0$. However, the characteristics are shown in Fig. 3.3, suggesting that u may change much more rapidly in the η direction than in the ξ direction. In fact, if we neglect ξ derivatives altogether in (3.38), we obtain the approximation[34]

$$u \approx f(\xi) - \epsilon^2 g(\xi) e^{-\eta/\epsilon^2},$$

which suggests that, when η is much bigger than ϵ, the solution has 'forgotten' about $g(\xi)$ and only remembers the data $u = f(\xi)$. We will verify that $u = f$ is

[33] Recall that we have seen in §§2.4.2 and 2.4.3 that systems with a double characteristic can exhibit the structure of hyperbolic systems.

[34] Technically, there is a 'boundary layer' on $\eta = 0$.

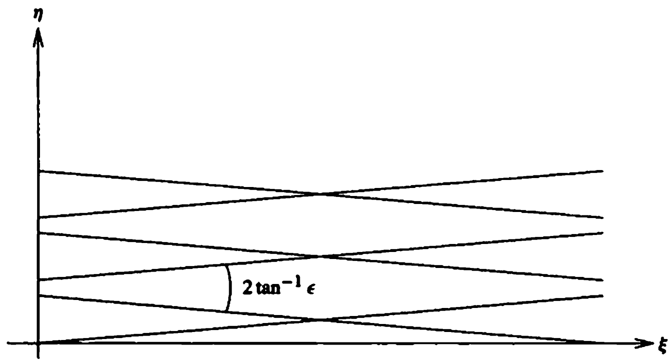


Fig. 3.3 Characteristics for the telegraph equation.

in fact correct Cauchy data for the diffusion equation in $\eta \geqslant 0$ when we come to Chapter 6, although, as example (3.2) shows, it would not be for the 'backward' diffusion equation $\partial^2 u/\partial\xi^2 = -\partial u/\partial\eta$ in $\eta \geqslant 0$.

3.5 Some general remarks

Quasilinear equations

None of the classifications described above can be carried through with any precision when any of the coefficients a, b or c in (3.16) depends on u and/or its first derivatives. Unless we are so extremely lucky that (3.20) happens to be integrable, the best we can do is to 'freeze' these coefficients at some point or in some region of the (x, y) plane at which we presume we know their values reasonably accurately, and then proceed by regarding (3.16) as an equation with constant coefficients. In fact, as already remarked, we implicitly adopted this strategy when we set up the models (3.13) for transonic flow and (3.15) for a glass sheet. Nevertheless, the information thus derived can be extremely valuable because not only does it reveal the range of behaviour that the local solution might display, but also, for hyperbolic problems, it provides the basis for an approximate iteration scheme on the lines proposed in Figs 2.6 and 2.9. The dangers inherent in basing one's ideas on 'localised' approximations come not so much from misinterpreting the reality or otherwise of the characteristics as in not noticing how nonlinearity may cause singularities to develop. Indeed, we have seen in §1.4 how this can happen even for first-order equations.

Goursat problems

In §3.2 we have only considered Cauchy problems, which, as we know, are well posed for hyperbolic equations. However, some practical problems lead to hyperbolic equations for which, instead of Cauchy data such that u and $\partial u/\partial n$ are given on a non-characteristic Γ, only u is prescribed on such a curve Γ. However, to compensate, u is also given on a characteristic, say $\xi = \xi_0$, in such a way as not to violate any of the equations that are known to hold along that characteristic

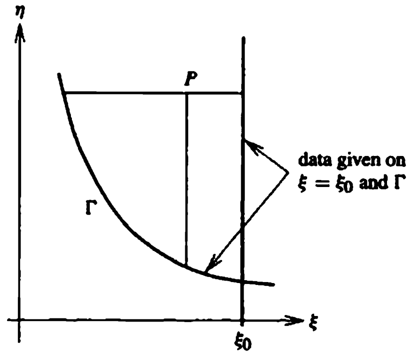


Fig. 3.4 Characteristics through P for the Goursat problem.

(Fig. 3.4). This is called a *Goursat* problem. It is easy to guess that it is well posed by simply counting the number of pieces of information carried along the characteristics passing through a typical point P, as we did in Figs 2.6 and 2.7. For example, if we know the two independent Riemann invariants explicitly, the data on Γ determines one functional relation between them and data on $\xi = \xi_0$ determines another.[35] We will return to this in Chapter 4.

Subcharacteristics

There is a piece of jargon that is sometimes used when approximations are sought for the solutions of second-order equations on the basis that some or all of the principal part is multiplied by a small parameter, as in the telegraph equation (3.38): such problems fall within the realm of ***singular perturbation*** theory and are well described in texts such as [3, 22, 25]. In particular, one phenomenon that emerges is that, whenever the principal part is neglected entirely, we are generally left with a scalar first-order equation, and such an equation always has a one-parameter family of real characteristics. These curves are sometimes called ***subcharacteristics*** and they can clearly carry much important information about the full solution. Unfortunately, they often behave in a troublesome manner in the vicinity of boundaries, particularly when the principal part of the equation is elliptic or when they 'graze' the boundary. Examples of both types of behaviour are given in Exercises 3.9 and 3.10.

More independent variables

When there are three or more independent variables, there is not much that can be said here without getting into the discussion that we gave at the end of Chapter 2. However, we can consider the generalisations of (3.6) and (3.7) to m dimensions for a constant-coefficient equation of the form

[35] A particularly interesting example of this situation concerns the Goursat problem for the Tricomi equation (3.13) when Γ is a curve that crosses the sonic line $U = 0$.

$$\frac{\partial^2 u}{\partial y^2} = \sum_{i=1}^{m-1} c_i \frac{\partial^2 u}{\partial x_i^2}.$$

It is only when $c_i > 0$ for all i that the Fourier series solution with Cauchy data

$$u|_{y=0} = \sum_{n_1,\dots,n_{m-1}=0}^{\infty} a_{n_1\dots n_{m-1}} \cos(n_1 x_1)\cdots\cos(n_{m-1}x_{m-1}), \qquad \left.\frac{\partial u}{\partial y}\right|_{y=0} = 0,$$

can be written as

$$u = \sum_{n_1,\dots,n_{m-1}=0}^{\infty} a_{n_1\dots n_{m-1}} \cos(ny)\cos(n_1 x_1)\cdots\cos(n_{m-1}x_{m-1}),$$

with $n^2 = \sum_{i=1}^{m-1} c_i n_i^2$. This can be put more elegantly by saying that the solution grows exponentially in y for most choices of the real 'wavenumbers' n_i unless the real quadratic form

$$\sum_{i=1}^{m-1} c_i \left(\frac{\partial \phi}{\partial x_i}\right)^2 - \left(\frac{\partial \phi}{\partial y}\right)^2$$

(i.e. the generalisation of the left-hand side of (3.19)) has rank m and signature $m-1$. This characterisation of the quadratic form, which is invariant under the kind of change of variables we considered in §3.4, is the basis of the generalised definition of hyperbolicity for second-order equations with an arbitrary number of independent variables. Unfortunately, it takes as many adjectives to describe non-hyperbolic equations as it does to encompass all the rank and signature possibilities of the relevant quadratic form. However, when the rank and signature are both m the word elliptic is usually used. As explained in §2.6, characteristics of hyperbolic equations are now manifolds, of dimension $m-1$, which locally near any point (y, x_i) take the form of cones of one sheet enclosing the 'time-like' y direction.

Exercises

3.1. Suppose that a, b and c are constant and consider the two equations

$$L_1 u = \frac{\partial^2 u}{\partial \xi \partial \eta} - a\frac{\partial u}{\partial \xi} - b\frac{\partial u}{\partial \eta} - cu = 0, \quad L_2 u = \frac{\partial^2 u}{\partial \xi^2} - a\frac{\partial u}{\partial \xi} - b\frac{\partial u}{\partial \eta} - cu = 0.$$

Write down the general solution in terms of two independent arbitrary functions of ξ and η

(a) for L_1, when $c = 0$ and either a or b vanishes;

(b) for L_2, when b vanishes.

Separate the variables to show that in each case solutions can be found in the form of exponentials in ξ and η.

3.2. Suppose $D_j = a_j\,\partial/\partial x + b_j\,\partial/\partial y + c_j$, $j = 1, \ldots, 4$, where a_j, b_j and c_j are constant, and

$$D_1 u + D_2 v = 0,$$
$$D_3 u + D_4 v = 0.$$

Show that u and v satisfy scalar second-order equations whose type depends on the roots of

$$\det\left(\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} - \lambda\begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}\right) = 0.$$

3.3. Show that the equations

$$\text{(i)}\ \frac{\partial^2 u}{\partial x^2} = 0, \quad \text{(ii)}\ \frac{\partial^2 u}{\partial x^2} = u, \quad \text{(iii)}\ \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial x}, \quad \text{(iv)}\ \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial y}$$

can be written as first-order systems for $\mathbf{u} = (u, v)^{\mathsf{T}}$, where $v = \partial u/\partial x$, in the form

$$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial y} = \mathbf{c},$$

where for cases (i)–(iii)

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and for case (iv)

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}.$$

Show further that, when λ is such that $\det(\mathbf{B} - \lambda\mathbf{A}) = 0$, then $\mathbf{B} - \lambda\mathbf{A}$ has two linearly independent left eigenvectors in cases (i)–(iii), but only one in case (iv). Show also that (i)–(iii) can usually be made to satisfy Cauchy data, but that (iv) cannot.

3.4. By transforming to canonical form, show that the general solution of

$$\frac{\partial^2 u}{\partial x^2} + \frac{10}{3}\frac{\partial^2 u}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} + \sin(x + y) = 0$$

is $u = f(x - 3y) + g(y - 3x) + \frac{3}{16}\sin(x + y)$, where f and g are arbitrary.

3.5. Show that the Tricomi-type equations

(a) $$x\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

(b) $$-y\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

(c) $$-x^2\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

are hyperbolic in $x < 0, y > 0$ and $x \neq 0$, respectively, and that characteristic coordinates are

(a) $$\xi = y + 2(-x)^{1/2}, \quad \eta = y - 2(-x)^{1/2},$$

(b) $$\xi = 3x + 2y^{3/2}, \quad \eta = 3x - 2y^{3/2},$$

(c) $$\xi = y + \log|x|, \quad \eta = y - \log|x|.$$

Sketch the characteristics in each case and show that the corresponding canonical forms are

(a) $$\frac{\partial^2 u}{\partial\xi\partial\eta} = \left(\frac{\partial u}{\partial\xi} - \frac{\partial u}{\partial\eta}\right)\Big/ 2(\eta - \xi),$$

(b) $$\frac{\partial^2 u}{\partial\xi\partial\eta} = \left(\frac{\partial u}{\partial\xi} - \frac{\partial u}{\partial\eta}\right)\Big/ 6(\xi - \eta),$$

(c) $$\frac{\partial^2 u}{\partial\xi\partial\eta} = \left(\frac{\partial u}{\partial\eta} - \frac{\partial u}{\partial\xi}\right)\Big/ 4.$$

3.6. Show that, if

$$(x + \alpha y)\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

where $\alpha \neq 0$, the characteristics are given by

$$(z \pm \alpha)^2 \mp 4\alpha z + 2\alpha^2 \log(z \pm \alpha) + x = \text{constant},$$

where $z^2 = x + \alpha y$. Show that, for small z, the characteristics through the origin are

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

What happens if $\alpha = 0$?

3.7. Show that the equation

$$2x^2\frac{\partial^2 u}{\partial x^2} + 5xy\frac{\partial^2 u}{\partial x\partial y} + 2y^2\frac{\partial^2 u}{\partial y^2} + 8x\frac{\partial u}{\partial x} + 5y\frac{\partial u}{\partial y} = 0$$

is hyperbolic and that characteristic coordinates are $\xi = x^2/y$ and $\eta = y^2/x$. Giving yourself lots of time, show that the canonical form is

$$\eta\frac{\partial^2 u}{\partial\xi\partial\eta} + \frac{\partial u}{\partial\xi} = 0,$$

and hence that the general solution is $u(\xi, \eta) = f(\xi)/\eta + g(\eta)$, where f and g are arbitrary. Hence, or otherwise, when $u(1, y) = y^2$ and $\partial u/\partial x\,(1, y) = 1$, show that

$$u(x, y) = \frac{1}{6}\left(2 - \frac{2}{x^3} + \frac{7y^2}{x} - \frac{y^2}{x^7}\right).$$

3.8. Show that, when $u(x,0) = u_0(x)$ and $\partial u/\partial t\,(x,0) = v_0(x)$, the d'Alembert representation (3.31) gives

$$u(x,t) = \frac{1}{2}\big(u_0(x-a_0t) + u_0(x+a_0t)\big) + \frac{1}{2a_0}\int_{x-a_0t}^{x+a_0t} v_0(s)\,ds.$$

3.9. Show that constants a and b can be chosen so that $u = ay + bye^{-x/\epsilon}$ satisfies

$$\epsilon\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\partial u}{\partial x} = 0 \quad \text{for } -1 < x < 1,$$

with $u(1,y) = y$ and $u(-1,y) = -y$. Check that, as $\epsilon \to 0+$, u becomes constant on the subcharacteristics $y =$ constant and satisfies one boundary condition, but not the other. What happens when $\epsilon \to 0-$?

3.10. Suppose that $u(x,y,t)$ satisfies the *elliptic* equation

$$\frac{\partial^2 u}{\partial y^2} + \epsilon^2\frac{\partial^2 u}{\partial x^2} = 0 \quad \text{for } 0 < y < 1,$$

with

$$\frac{\partial u}{\partial y}(x,0,t) = 0, \qquad \frac{\partial u}{\partial y}(x,1,t) + \epsilon^2\frac{\partial^2 u}{\partial t^2}(x,1,t) = 0.$$

Show that a formal power series in which

$$u = u_0(x,y,t) + \epsilon^2 u_1(x,y,t) + \cdots$$

satisfies the equation and boundary condition up to terms of $O(\epsilon^2)$ if

$$u_0 = u_0(x,t),$$

where u_0 satisfies the *hyperbolic* equation

$$\frac{\partial^2 u_0}{\partial x^2} - \frac{\partial^2 u_0}{\partial t^2} = 0.$$

This example, which is a simple model for the tides, shows that the solution of an elliptic equation can sometimes be consistently approximated by that of a hyperbolic equation.

4

Hyperbolic equations

4.1 Introduction

Hyperbolic equations are the easiest scalar second-order equations to classify from the point of view of the Cauchy problem. They occur commonly in practical applications, as is evident from studying the models of Chapter 2. Take, for example, fluid dynamics. We have already remarked in Chapter 3 that a large class of steady two-dimensional supersonic gas flows can be modelled by the hyperbolic equation (3.11), $(1-U^2/a_0^2)\partial^2\phi/\partial x^2+\partial^2\phi/\partial y^2=0$. However, the requisite cross-differentiation is not possible for evolutionary models such as the shallow water model (2.1) and (2.2), or the unsteady gas dynamics model (2.3) and (2.4), except in the frequently occurring *acoustic limit* in which the fluid is nearly in a state of rest or uniform motion. Then, a linearisation procedure can be carried out as on p. 65. For example, when we assume that u and $h-h_0$ in (2.1) and (2.2), or u, $\rho-\rho_0$ and $p-p_0$ in (2.3) and (2.4), are small enough for their squares to be negligible, we obtain the hyperbolic *one-dimensional wave equation*

$$\frac{\partial^2\phi}{\partial t^2}=a_0^2\frac{\partial^2\phi}{\partial x^2}, \tag{4.1}$$

where $a_0^2=gh_0$ or $a_0^2=\gamma p_0/\rho_0$, respectively. Here ϕ is any one of the variables u or $h-h_0$ in (2.1) and (2.2), or u, $\rho-\rho_0$ or $p-p_0$ in (2.3) and (2.4), the remaining variables being related to ϕ by simple linear transformations. Equally, the regenerator and fluidised bed models (2.17)–(2.20) can be cross-differentiated to give a constant-coefficient equation of the form

$$\left(\frac{\partial}{\partial t}+a\frac{\partial}{\partial x}+b\right)\left(\frac{\partial}{\partial t}+c\frac{\partial}{\partial x}+d\right)u=eu, \tag{4.2}$$

which is clearly hyperbolic and is a version of the telegraph equation mentioned in §3.4.3; to see this set $u=e^{\alpha x+\beta t}v$ for suitable α and β.

Multi-space-dimensional versions of such linear scalar second-order equations are even more relevant for practical problems. Indeed, much of linear acoustics is governed by

$$\frac{\partial^2 u}{\partial t^2}=a_0^2\nabla^2 u \tag{4.3}$$

in two or three space dimensions,[36] and we will discuss how the solutions of such hyperbolic equations depend on dimensionality in §4.6. Still more interesting is the appearance of vector versions of (4.3), which should really be classified using the methods of Chapter 2. However, a generalisation of the discussion of §3.4 will give us enough of a basis to be able to discuss systems such as Maxwell's equations and the equations of linear elasticity in §4.7.

Our first task in this chapter is to see how the theory for linear scalar second-order equations in two independent variables can be put to good use. Since we know that the Cauchy problem for such equations is well posed, we can immediately set about finding representations of solutions, confident that such representations make sense and depend continuously on the data. This is the first time in the book that a general strategy for representing solutions is presented, and it is not the last, so a careful preamble about the possible procedures and methodology is given in §4.2. Moreover, there is one analytical technique, that of eigenfunction expansions, which we review in §4.4 because it is of such importance in this and subsequent chapters. Most of the rest of this chapter is devoted to the explicit solution of linear equations, and we leave those few representations that are available for the solution of nonlinear hyperbolic problems to the brief §4.8.

4.2 Linear equations: the solution to the Cauchy problem

We begin by considering linear equations that have been transformed to characteristic variables (x, y), so that we have to solve

$$\mathcal{L}u = \frac{\partial^2 u}{\partial x \partial y} + p\frac{\partial u}{\partial x} + q\frac{\partial u}{\partial y} + ru = f, \tag{4.4}$$

where p, q, r and f are functions of x and y, but not of u or its derivatives. We take u and $\partial u/\partial n$ (or, equivalently up to a constant, $\partial u/\partial x$ and $\partial u/\partial y$) to be prescribed on some open curve Γ that is nowhere parallel to the characteristics, which are the x and y axes. For definiteness we only look for the solution on one side of this curve, say y increasing, as in Fig. 4.1.

The prescription for the representation for the solution of this general Cauchy problem can be presented at two different levels and we leave it to the reader to decide which route to take; they converge at the end of §4.2.2.

4.2.1 An *ad hoc* approach to Riemann functions

Motivated by §§3.3 and 3.4.1, we ask whether the d'Alembert formula (3.31) could not be generalised to apply to (4.4). Now the key step in deriving (3.31) was direct integration with respect to first x and then y, which is one of the many ways of evaluating a double integral over the region D of Fig. 4.1. In the light of

[36] Here and henceforth we will use the applied mathematician's notation

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

where x, y, z denote spatial coordinates.

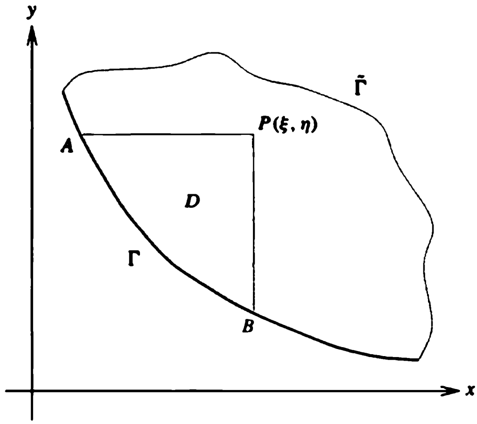


Fig. 4.1 The Cauchy problem for $u(\xi,\eta)$.

this observation, and remembering our discussion of transpose (adjoint) matrices on p. 43, we consider defining an operator $\mathcal{L}^*$, called the *adjoint* of $\mathcal{L}$, such that $u\mathcal{L}v - v\mathcal{L}^*u$ is a divergence, i.e. it is of the form $\partial P/\partial x + \partial Q/\partial y$, no matter what the functions u and v are. This will enable its double integral over D to be replaced by a line integral around the boundary ∂D. Inspection of (4.4) shows that the only possibility is to define

$$\mathcal{L}^*v = \frac{\partial^2 v}{\partial x \partial y} - \frac{\partial}{\partial x}(pv) - \frac{\partial}{\partial y}(qv) + rv, \tag{4.5}$$

so that we may use Green's theorem to obtain

$$\iint_D (v\mathcal{L}u - u\mathcal{L}^*v)\,\mathrm{d}x\,\mathrm{d}y = \oint_{\partial D} \left(puv + v\frac{\partial u}{\partial y}\right)\mathrm{d}y + \left(u\frac{\partial v}{\partial x} - quv\right)\mathrm{d}x. \tag{4.6}$$

Now we want to choose v so that this formula can be used to tell us the value $u(\xi,\eta)$ of the solution at P. We see that we can remove all the terms whose values we do not know if v satisfies

$$\begin{aligned} \mathcal{L}^*v &= 0 \quad \text{in } D, \\ \frac{\partial v}{\partial x} &= qv \quad \text{on } AP, \text{ on which } y = \eta, \\ \frac{\partial v}{\partial y} &= pv \quad \text{on } BP, \text{ on which } x = \xi, \end{aligned} \tag{4.7}$$

because the integrals along AP and BP can then be evaluated explicitly, integrating by parts where necessary. This gives

$$[uv]_B^P = \iint_D vf\,\mathrm{d}x\,\mathrm{d}y - \int_A^B \left(puv + v\frac{\partial u}{\partial y}\right)\mathrm{d}y + \left(u\frac{\partial v}{\partial x} - quv\right)\mathrm{d}x. \tag{4.8}$$

We could have equally well found $[uv]_A^P$, but the two results would have been simply related by an integration by parts from A to B. Finally, we note that v

is undefined to within a multiplicative constant so that, if v exists, it can be set equal to unity at P.

The function v is called the *Riemann function* for our problem and, if we can convince ourselves that it exists, (4.8) gives us our desired representation of the solution of the Cauchy problem. In fact, existence is no problem when (4.5) can be formally converted to a linear Volterra integral equation; this happens, for example, if $p = q = 0$, when

$$v(x,y) = 1 - \int_\eta^y \int_\xi^x r(x',y')v(x',y')\,\mathrm{d}x'\,\mathrm{d}y';$$

the convergence of a sequence satisfying '$v_{n+1} = 1 - \iint rv_n$' to a unique $v(x,y)$ is easy to establish, as in the proof of Picard's theorem for ordinary differential equations.

Even in the absence of any explicit formula for the Riemann function v, our basic result (4.8) contains a wealth of valuable information about hyperbolic equations, to which we will return at the end of this section. However, we first present an alternative, more complicated, but more systematic, story of the Riemann function which may be skipped by those anxious to get to explicit situations as quickly as possible. We nonetheless urge those who wish to see a coherence between this chapter and its successors to read on.

4.2.2 The rationale for Riemann functions

We begin this rather lengthy tale by recalling the Fredholm Alternative that we encountered in §2.2 for the matrix equation

$$\mathcal{A}\mathbf{x} = \mathbf{b}, \tag{4.9}$$

where $\mathbf{x}$ and $\mathbf{b}$ are column vectors, and $\mathcal{A}$ is an $n \times n$ matrix. If the solution exists, we can simply write it as $\mathbf{x} = \mathcal{A}^{-1}\mathbf{b}$. However, we can write this in an especially convenient way if we observe that, if $\mathbf{y}_k$ is a vector such that[37]

$$\mathbf{y}_k^\mathsf{T}\mathcal{A} = \mathbf{e}_k^\mathsf{T}, \quad \text{i.e.} \quad \mathcal{A}^\mathsf{T}\mathbf{y}_k = \mathbf{e}_k, \tag{4.10}$$

where $\mathbf{e}_k$ is the kth standard basis vector, then the kth component of $\mathbf{x}$, which is equal to $\mathbf{e}_k^\mathsf{T}\mathbf{x}$, can be written

$$(\mathbf{y}_k^\mathsf{T}\mathcal{A})\mathbf{x} = \mathbf{y}_k^\mathsf{T}(\mathcal{A}\mathbf{x}) = \mathbf{y}_k^\mathsf{T}\mathbf{b}.$$

Thus we have found the kth component of $\mathbf{x}$ simply by taking the 'inner product' of (4.9) with $\mathbf{y}_k$ and of (4.10) with $\mathbf{x}$ and subtracting; all we needed was the identity

$$\mathbf{x}^\mathsf{T}\mathcal{A}^\mathsf{T}\mathbf{y}_k = \mathbf{y}_k^\mathsf{T}\mathcal{A}\mathbf{x}.$$

This offers us a layout that can be used repeatedly for defining the inverse of all kinds of linear differential operators. For example, suppose we consider the initial value problem for a second-order linear ordinary differential equation in the form

[37] Of course, $\mathbf{y}_k^\mathsf{T}$ is just the kth row of $\mathcal{A}^{-1}$.

$$\mathcal{L}u = f(x), \quad \mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}}{\mathrm{d}x} + q(x) \quad \text{for } x > 0, \tag{4.11}$$

with $u(0) = \mathrm{d}u/\mathrm{d}x\,(0) = 0$ without loss of generality. We exploit the argument above by identifying $\mathcal{L}$ with $\mathcal{A}$, u with $\mathbf{x}$ and f with $\mathbf{b}$. We now try to define $\mathcal{L}^*$, δ, R and an inner product $(\cdot\,,\cdot)$ so that we can draw the analogies

$$\begin{array}{rcc}
\mathcal{A}\mathbf{x} = \mathbf{b} & \leftrightarrow & \mathcal{L}u(x) = f(x), \quad u(0) = \frac{\mathrm{d}u}{\mathrm{d}x}(0) = 0; \\
\mathcal{A}^{\mathsf{T}}\mathbf{y}_k = \mathbf{e}_k & \leftrightarrow & \mathcal{L}^* R(x,\xi) = \delta(x-\xi); \\
\mathbf{y}_k^{\mathsf{T}}\mathcal{A}\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathcal{A}^{\mathsf{T}}\mathbf{y}_k & \leftrightarrow & \big(R(x,\xi), \mathcal{L}u(x)\big) = \big(u(x), \mathcal{L}^* R(x,\xi)\big); \\
\mathbf{y}_k^{\mathsf{T}}\mathbf{b} = \mathbf{e}_k^{\mathsf{T}}\mathbf{x} & \leftrightarrow & \big(R(x,\xi), f(x)\big) = \big(\delta(x-\xi), u(x)\big).
\end{array}$$

The last line in the left-hand column gives $\mathbf{x}$, component by component, and the last line in the right-hand column is supposed to give $u(\xi)$ at all possible values of ξ. These analogies will work if we proceed as follows.

1. We define $\big(u(x), v(x)\big) = \int_0^X u(x)v(x)\,\mathrm{d}x$ for some suitable $X > \xi$. The motivation for this is as follows: if we divide the range of integration into N subintervals of width h, and approximate $u(x)$ and $v(x)$ by step functions on each subinterval, the functions are represented by vectors of values $\mathbf{u} = (u_1, \ldots, u_N)$ and $\mathbf{v} = (v_1, \ldots, v_N)$. The integral of uv is approximated by $\sum u_i v_i h = \mathbf{u}\cdot\mathbf{v}\,h$ which, under suitable conditions, becomes $\int uv\,\mathrm{d}x$ as $N \to \infty$.
2. We define
$$\mathcal{L}^* R = \frac{\mathrm{d}^2 R}{\mathrm{d}x^2} - \frac{\mathrm{d}}{\mathrm{d}x}(pR) + qR,$$
which makes $u\mathcal{L}^* R - R\mathcal{L}u$ into an exact differential, and in addition we specify that $R(X) = \mathrm{d}R/\mathrm{d}x\,(X) = 0$.
3. We define $\delta(x-\xi)$ to be the so-called *delta function*, i.e. the limit of a sequence of well-behaved functions that tend to zero except for very small values of $x-\xi$, but whose integral over any interval containing $x = \xi$ is unity. The sequence could, for example, be defined by
$$\delta(x) = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon\sqrt{\pi}} \mathrm{e}^{-x^2/\epsilon^2}. \tag{4.12}$$
By taking limits of integrals of members of this approximating sequence, the vital result $\big(\delta(x-\xi), u(x)\big) = u(\xi)$ in the last line of the analogy can be justified as long as $X > \xi$ and, strikingly, it can be shown to follow from limits of any reasonable sequence of approximating sequences like (4.12).

This last idea is one of the principal motivations for the axiomatic definition of *distributions* or *generalised functions* [37, 42]. In this approach all the 'epsilonology' is swept away by defining the delta function and its relatives as linear functionals that map a class of suitably smooth *test functions* to the real numbers.[38] We define

[38] The whole philosophy is closely related to that of the theory of weak solutions of hyperbolic equations in the presence of shock waves, as discussed in §1.7.

this map by the inner product $(\cdot\,,\cdot)$; thus, since the inner product of $\delta(x)$ and any well-behaved test function $\psi(x)$ is $\psi(0)$, we could define δ axiomatically by the statement

$$\delta(x) : \psi \mapsto \psi(0).$$

In this scenario, any well-behaved function $f(x)$ would be defined as

$$f(x) : \psi \mapsto \int_{-\infty}^{\infty} f(x)\psi(x)\,\mathrm{d}x$$

for all test functions ψ, assuming of course that the test functions are such as to allow the integral to exist. We can go further and define a calculus of distributions motivated by the formula for integration by parts, which says that

$$\int_{-\infty}^{\infty} f'(x)\psi(x)\,\mathrm{d}x = -\int_{-\infty}^{\infty} f(x)\psi'(x)\,\mathrm{d}x,$$

provided that ψ vanishes sufficiently rapidly at $x = \pm\infty$. Thus we define $\delta'(x)$ by the rule

$$\int_{-\infty}^{\infty} \delta'(x)\psi(x)\,\mathrm{d}x = -\int_{-\infty}^{\infty} \delta(x)\psi'(x)\,\mathrm{d}x,$$

so that

$$\delta'(x) : \psi \mapsto -\psi'(0);$$

the intuitively obvious statements

$$\int_{-\infty}^{x} \delta(\xi)\,\mathrm{d}\xi = H(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0, \end{cases} \tag{4.13}$$

where $H(x)$ is called the ***Heaviside function***, and its consequence

$$H'(x) = \delta(x)$$

are in accordance with this rule. It can be shown that the delta function, like all distributions, satisfies the usual rules of calculus, especially that of integration by parts for the product of a distribution and a smooth function. Moreover, it is a simple matter to define distributions with, say, two independent variables; for example, we could either consider $\delta(x)\delta(y)$ to be defined by[39]

$$\delta(x)\delta(y) = \lim_{\epsilon\to 0} \frac{1}{\pi\epsilon^2}\mathrm{e}^{-(x^2+y^2)/\epsilon^2},$$

or as the functional that takes a test function $\psi(x, y)$ to $\psi(0, 0)$.

[39]It is interesting and important to show that the interpretation of $\delta(x)\delta(y)$ is independent of any local change of coordinates that preserves area (see Exercise 4.1).

At the end of the day, this calculus of distributions, combined with the analogy above, leads to the solution of (4.11) in the form

$$\begin{aligned} u(\xi) &= \int_0^X u\mathcal{L}^* R \,\mathrm{d}x = \int_0^X u(x)\delta(x-\xi)\,\mathrm{d}x \\ &= \int_0^X R\mathcal{L}u \,\mathrm{d}x \;= \int_0^X R(x,\xi)f(x)\,\mathrm{d}x. \end{aligned}$$

The heart of the matter is that Green's theorem (here, integration by parts) can be used to relate the integrals of $u\mathcal{L}^*R$ and $R\mathcal{L}u$ despite the fact that R has singularities and that $\mathcal{L}R$ is not even a function in the usual sense. Moreover, noting that R satisfies the linear homogeneous equation $\mathcal{L}^*R = 0$ in $\xi < x < X$, with $R = \mathrm{d}R/\mathrm{d}x = 0$ at $x = X$, we see that $R \equiv 0$ for $x > \xi$. Thus we simply have

$$u(\xi) = \int_0^{\xi} R(x,\xi)f(x)\,\mathrm{d}x.$$

As a matter of jargon, operators $\mathcal{L}$ such that $\mathcal{L} = \mathcal{L}^*$ are called *self-adjoint*, and problems for which $\mathcal{L} = \mathcal{L}^*$ and additionally R satisfies the same boundary conditions as those for u are called *self-adjoint problems*. Cauchy problems for hyperbolic equations do not fall into this category, even though $\mathcal{L}$ may equal $\mathcal{L}^*$, but certain problems for elliptic equations do, as we shall see in Chapter 5. In the same way that self-adjoint real matrices, i.e. symmetric ones, have real eigenvalues and orthonormal eigenvectors, so do self-adjoint real differential operators have real eigenvalues and orthonormal eigenfunctions.

The procedure (4.5)–(4.8) now becomes an obvious generalisation. We simply define

$$\mathcal{L}^*R(x,y;\xi,\eta) = \delta(x-\xi)\delta(y-\eta).$$

The nice thing about the layout on p. 97 is that, if we now take the boundary conditions on R as Cauchy data

$$R = \frac{\partial R}{\partial n} = 0$$

on any curve outside D (such as $\tilde{\Gamma}$ in Fig. 4.1), we immediately retrieve (4.8) formally with $v = R$, $\mathcal{L}^*$ as in (4.5) and the inner product

$$(u(x,y), v(x,y)) = \iint_D uv \,\mathrm{d}x\,\mathrm{d}y.$$

We also note that $R \equiv 0$ in the unshaded region between $\tilde{\Gamma}$ and D in Fig. 4.1, because R satisfies homogeneous Cauchy data on $\tilde{\Gamma}$.

However, the argument above raises a much more subtle question about the nature of the Riemann function. Since R describes the response of a hyperbolic equation to an 'impulse' at $x = \xi$, $y = \eta$, our general arguments about characteristics lead us to expect that R vanishes except in $x < \xi$, $y < \eta$, and suffers discontinuities of some sort on $x = \xi$, $y < \eta$, and on $y = \eta$, $x < \xi$ (i.e. BP and AP, respectively, in Fig. 4.1). But what can we say about these discontinuities? This is a difficult question whose answer demands that we first look at R in the vicinity of $x = \xi$, $y = \eta$. Here, we expect the largest term in $\mathcal{L}^* R$ to be the highest derivative $\partial^2 R/\partial x\,\partial y$ and, by direct integration and the use of (4.13),

$$\frac{\partial^2 R}{\partial x\,\partial y} = \delta(x-\xi)\delta(y-\eta) \quad \text{implies} \quad R = H(\xi - x)H(\eta - y);$$

that is, locally near $x = \xi$, $y = \eta$, R is unity in the quadrant $x < \xi$, $y < \eta$, and zero outside. Note that this confirms that, as we approach P from inside D, R tends to unity in accordance with the statement after (4.8).

Now, since R suffers jumps across $x = \xi$ and $y = \eta$ locally near P, we expect from arguments such as those in §1.6 that these jumps persist although they may change in strength, depending on the form of $\mathcal{L}$. To find these changes, say across $x = \xi$, we note that when we formally integrate $\mathcal{L}^* R = 0$ with respect to x across BP we find

$$\left[\frac{\partial R}{\partial y}\right]_{x=\xi-0}^{x=\xi+0} - \left[pR\right]_{x=\xi-0}^{x=\xi+0} = 0.$$

But $R \equiv 0$ on $x = \xi + 0$ so, as we approach $x = \xi$ from D,

$$\frac{\partial R}{\partial y} - pR = 0,$$

in accordance with (4.7). The same kind of argument applies on AP.

We thus have achieved a complete agreement between our constructive approach to the Riemann function and the *ad hoc* approach of §4.2.1. While §4.2.1 may seem the easier, we will find that the hard work of this section will pay handsome dividends in the remaining chapters and especially at the end of Chapter 9.

4.2.3 Implications of the Riemann function representation

From now on we will reserve the notation R for the Riemann function so that (4.8) becomes

$$u(P) - u(B) = \iint_D Rf\,\mathrm{d}x\,\mathrm{d}y - \int_A^B \left(puR + R\frac{\partial u}{\partial y}\right)\mathrm{d}y + \left(u\frac{\partial R}{\partial x} - quR\right)\mathrm{d}x. \tag{4.14}$$

We remember that we could have written down a similar formula for $u(P) - u(A)$ and used it with (4.14) to derive the d'Alembert formula (3.31). We cannot write down an explicit expression for R for the general hyperbolic equation (4.4), although some examples where this is possible are given below. Nevertheless, the existence of this function R, which is defined independently of the boundary data and only depends on the differential operator, gives valuable qualitative information about the solution of (4.4) which we summarise below.

1. The solution only depends on the Cauchy data between A and B, where the characteristics through P meet the boundary curve, and the values of f in D, the region between the two characteristics through P and the boundary curve. This region is called the ***domain of dependence*** of P, as shaded in Fig. 4.1.

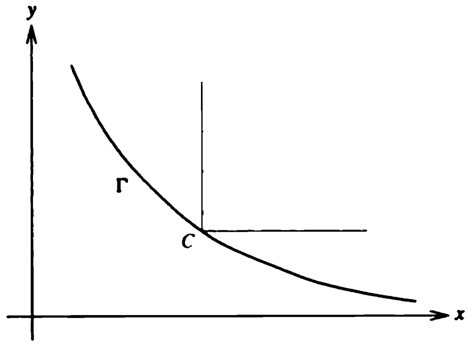


Fig. 4.2 Region of influence of C.

2. The Cauchy data at a point C on the boundary curve only influences the solution in the shaded quadrant in Fig. 4.2. This is called the *region of influence* of C.

   Both properties 1 and 2 are obvious from (4.14), and are special cases of the corresponding definitions in Chapter 2 for a hyperbolic system.

3. The Cauchy data on AB defines a unique solution in the triangle APB in Fig. 4.1, and corresponds to the domain of definition discussed in §2.4 for the first-order system. This is also easily obtained from (4.14), given that the Riemann function exists, by considering the difference of two possible solutions satisfying the same equation and boundary conditions.

4. The solution u at P depends continuously on f, p, q, r and on the boundary data, in the sense that a small change in the values of $\partial u/\partial x$ or $\partial u/\partial y$ on the boundary curve Γ, or in the shape of Γ, results in a small change in the value of u at P. This property, already anticipated in Chapter 2, may be obtained from (4.14) by use of suitable inequalities, given that f, p, q and r are smooth functions, and it confirms that the Cauchy problem is well posed.

5. Discontinuities in the form of jumps in the second derivatives of u in the boundary data propagate along the characteristics through the point of discontinuity. It is not, however, obvious how jump discontinuities in the boundary values of the first derivatives of u or in u itself propagate along the characteristics through the point of discontinuity and this cannot be discussed without the introduction of a weak formulation. This will be done for the wave equation in the next section.

Almost the only useful explicit forms for Riemann functions are for the wave equation and the telegraph equation.

Example 4.1 (*The wave equation*) The canonical form for the wave equation (4.1) is

$$\mathcal{L}u = \frac{\partial^2 u}{\partial x \partial y} = f.$$

Then the adjoint equation is

$$\mathcal{L}^* R = \frac{\partial^2 R}{\partial x \partial y} = 0,$$

and $\mathcal{L}$ is self-adjoint. Also $R = 1$ on $y = \eta$ and $x = \xi$, so that $R \equiv 1$ (meaning, of course, that $R = H(\xi - x)H(\eta - y)$ in the whole plane, where $H(\cdot)$ is the Heaviside function); this was clearly going to happen the moment we compared (4.14) with the d'Alembert representation (3.31).

Example 4.2 (*The telegraph equation*) In canonical form the telegraph equation is

$$\mathcal{L}u = \frac{\partial^2 u}{\partial x \partial y} + u = f.$$

As suggested in §4.1, the first-order derivatives in (4.2) can be eliminated by simply multiplying u by an exponential function of x and t. The operator is again self-adjoint and $\partial^2 R/\partial x \partial y + R = 0$ with

$$R = 1 \quad \text{on } x = \xi, \ y \leqslant \eta, \text{ and on } y = \eta, \ x \leqslant \xi.$$

We can change the origin to (ξ, η) and note the 'symmetry' that, if $R = F(x - \xi, y - \eta)$ is a Riemann function, then so is $F\left(\lambda(x - \xi), \lambda^{-1}(y - \eta)\right)$, where λ is any constant. Thus R is only a function of the *similarity variable*[40] $(x - \xi)(y - \eta) = s$, say, and the manipulation is easier if we write this as $F(2\sqrt{s})$. Then F satisfies

$$\frac{\mathrm{d}^2 F}{\mathrm{d}s^2} + \frac{1}{s}\frac{\mathrm{d}F}{\mathrm{d}s} + F = 0 \qquad \text{for } s > 0,$$

and $F(0) = 1$. Hence $F(s) = J_0(s)$, the Bessel function of the first kind and zero order, and the Riemann function is

$$R = J_0\left(2\sqrt{(\xi - x)(\eta - y)}\right); \tag{4.15}$$

again, this needs to be multiplied by $H(\xi - x)H(\eta - y)$ to make it valid in the whole plane. A more direct, but more tedious, derivation of (4.15) is given in §4.5.

4.3 Non-Cauchy data for the wave equation

In practice, the data that we wish to impose on scalar hyperbolic equations may be different from Cauchy data prescribed, say, at an initial instant of time. In particular, there may be boundary conditions to be imposed as well as initial conditions, and there may be singularities in the data that would invalidate the derivation of the Riemann function representation (4.14). We have already encountered one non-Cauchy problem, namely the Goursat problem, in §3.5. Another frequently occurring case is that of the 'initial–boundary' value problem for the wave equation (4.1) when, say, we are modelling waves on a string of finite length.

[40]We will say more about this kind of situation in Chapter 6.

We begin by recalling our demonstration in the previous section that, given the function u and its normal derivative on an initial curve Γ which is nowhere tangent to either characteristic, then there is a unique solution for the linear hyperbolic problem. Moreover, in the first-order scalar case, any tangency between the boundary curve Γ and a characteristic meant that the problem could easily be ill-posed. Hence, an obvious question to ask is what kinds of data are possible for a second-order hyperbolic equation on a boundary curve Γ which touches a characteristic at a point. The initial–boundary value problem for (4.1), in the form

$$\frac{\partial^2 u}{\partial t^2} = a_0^2 \frac{\partial^2 u}{\partial x^2},$$

with Cauchy data given on $t = 0$, $x > 0$ and further boundary data given on $x = 0$, $t > 0$ (Fig. 4.3), is an extreme case; if the corner in this boundary curve was smoothed, there would be one point of tangency with a characteristic $x + a_0 t =$ constant.

This example is particularly convenient because the wave equation has the useful property that the Riemann invariants $\partial u/\partial t \pm a_0 \partial u/\partial x$ are constant on the characteristics $x \pm a_0 t =$ constant, respectively. Thus the data at $t = 0$ immediately allows us to find the solution at points $P = (\xi, \tau)$ such that $\xi > a_0 \tau$ (see Fig. 4.3). Moreover, if the solution u is required to be continuous, together with its first derivatives, across the characteristic $x = a_0 t$, then, on each 'negative' characteristic (on which $dx/dt = -a_0$) entering $0 < x < a_0 t$, the Riemann invariant $a_0 \partial u/\partial x + \partial u/\partial t$ is given by the values of u in $x > a_0 t$. Since each of these characteristics meets the boundary $x = 0$, $t > 0$, it is not in general possible to specify both $\partial u/\partial x$ and $\partial u/\partial t$ there. In fact, only one relation between them can be given, and the general allowable linear boundary condition has the form of a scalar specification

$$\alpha \frac{\partial u}{\partial x} + \beta \frac{\partial u}{\partial t} + \gamma u = \delta, \tag{4.16}$$

where the coefficients must be functions of t such that $\alpha - a_0 \beta$ and γ do not simultaneously vanish for any t. If, and only if, this last condition is satisfied, we

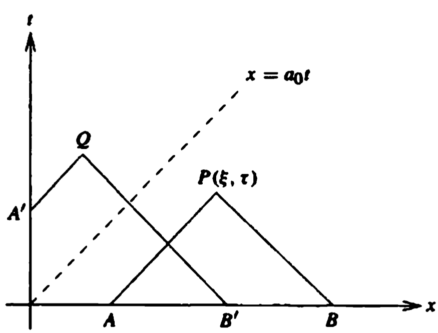


Fig. 4.3 Initial–boundary value problem for the wave equation.

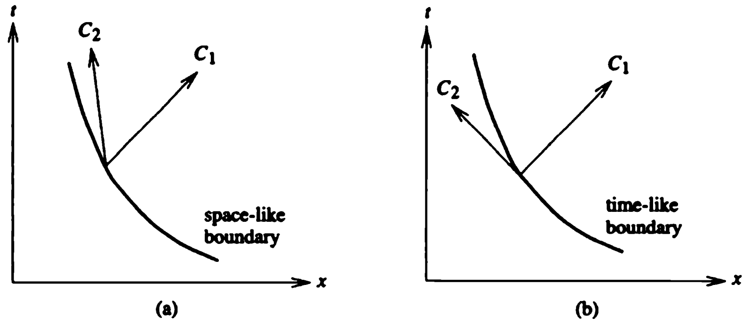


Fig. 4.4 Space- and time-like boundaries.

can determine $\partial u/\partial x$ and $\partial u/\partial t$ on $x = 0$, $t > 0$, and hence obtain the solution at the point Q in Fig. 4.3, from the Riemann invariants along QA' and QB'. Of course, the solution thus constructed usually has discontinuous second derivatives across the characteristic $x = a_0 t$. One way by which we could avoid such discontinuities would be to 'relax' the Cauchy data on $x = 0$. Indeed, if only u or $\partial u/\partial n$ was prescribed on $x = 0$ for all t and on $t = 0$ for all x, we could find a solution as smooth as the data, the Goursat problem of §3.5 being an example of this.

The discussion above illustrates the fact that if a boundary curve touches a characteristic then Cauchy data can only be given on one side of the point of tangency. The problem is, however, more or less symmetric in x and t so that it is apparently immaterial on which side the Cauchy data is posed; in our example this data could have been given on $x = 0$, $t > 0$ with only one condition on $t = 0$, $x > 0$. However, in this case the characteristics of slope $-a_0$ would be propagating the information $\partial u/\partial t + a_0 \partial u/\partial x =$ constant *backwards* in time, and such a model may be said to violate causality. Thus, if we associate with the characteristics a direction corresponding to time increasing, we are led to a characterisation of well-posed initial–boundary value problems. For such problems Cauchy data is prescribed on a curve whose slope $\mathrm{d}x/\mathrm{d}t$ is always greater than a_0 in modulus, and a single condition such as (4.16) is prescribed on a curve or curves whose slope is always less than a_0 in modulus.

Although with just one space variable there is no mathematical reason for introducing causality, the concept of a *time-like direction* can sometimes be introduced on physical grounds for other second-order hyperbolic equations. The discussion above shows that we need a time-like variable t and a *space-like* variable x, and two families of characteristics along each member of which a direction of increasing time can be determined.[41] Then, causality allows us to define a *space-like boundary* to be such that at all points both characteristics C_i in the positive time direction point toward the same side, as in Fig. 4.4(a), or, for example, on $t = 0$, $x > 0$ in

[41] Strictly speaking we should say that directions within a certain cone of directions defined by the eigenvalues of the relevant matrix are 'time-like' and that surfaces lying outside this cone are 'space-like' (see §2.6).

Fig. 4.3. Likewise, a *time-like boundary* separates the characteristics in the positive time direction, as in Fig. 4.4(b), or, for example, on $x = 0$, $t > 0$ in Fig. 4.3. We need Cauchy data, involving two boundary conditions, on a space-like boundary but only one condition is permissible on a time-like boundary, and, as in (4.16), this condition must not contradict the information propagating along one of the families of characteristics emanating from the Cauchy data.

*4.3.1 Strongly discontinuous boundary data

When the boundary data for a hyperbolic equation is not smooth, we may need to generalise our idea of what constitutes a solution. In §2.5, we introduced weak solutions for a first-order quasilinear hyperbolic system, and we now work through these ideas for the simpler case of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = a_0^2 \frac{\partial^2 u}{\partial x^2}.$$

If the first derivatives of u are discontinuous at just one point of $t = 0$, say $x = 0$, then the Riemann function representation defines a solution which has continuous derivatives except on $x = \pm a_0 t$. Following the ideas of weak solutions discussed in §1.7 for scalar first-order equations, we say that a weak solution is a function u which satisfies the identity

$$\iint_{t>0} u \left(\frac{\partial^2 \psi}{\partial t^2} - a_0^2 \frac{\partial^2 \psi}{\partial x^2} \right) \mathrm{d}x \, \mathrm{d}t = \int_{-\infty}^{\infty} \left(\psi \frac{\partial u}{\partial t} - u \frac{\partial \psi}{\partial t} \right)_{t=0} \mathrm{d}x, \tag{4.17}$$

where ψ is a twice-differentiable test function which vanishes suitably rapidly as $x \to \pm\infty$ and $t \to \infty$. If u is twice differentiable, integration of (4.17) by parts easily shows that it satisfies the wave equation everywhere in $t > 0$. However, we emphasise that (4.17) is a mathematical statement that has no physical foundation in any conservation law such as (1.24).

If, on the other hand, u is continuous but ∇u is discontinuous across a 'shock' curve C, we know that $[\partial u/\partial x]\,\mathrm{d}x + [\partial u/\partial t]\,\mathrm{d}t = 0$ across C and,[42] using an argument similar to that used in obtaining the Rankine–Hugoniot condition (1.27), we find

$$\int_C \psi \left(a_0^2 \left[\frac{\partial u}{\partial x} \right] \mathrm{d}t + \left[\frac{\partial u}{\partial t} \right] \mathrm{d}x \right) = 0,$$

for any test function ψ, so that

$$a_0^2 \left[\frac{\partial u}{\partial x} \right] \mathrm{d}t + \left[\frac{\partial u}{\partial t} \right] \mathrm{d}x = 0. \tag{4.18}$$

Thus, as expected, C is a characteristic through $x = t = 0$ with slope $\mathrm{d}x/\mathrm{d}t = \pm a_0$; across C, $a_0[\partial u/\partial x] \pm [\partial u/\partial t] = 0$ and we obtain the jump conditions necessary to define the solution uniquely in $-a_0 t < x < a_0 t$. For the wave equation, these

[42] We use the notation $[\cdot]$ in the same sense as in Chapter 1, to denote the jump from one side of the shock to the other.

jump conditions say that the Riemann invariants are continuous on the family of characteristics intersecting the line of discontinuity C, which is exactly what happened in the situation in Fig. 4.3. Because the appropriate Riemann invariants are continuous, the d'Alembert solution still provides us with the weak solutions of (4.1).

If u itself is discontinuous on the boundary, we may still use (4.18) to define a weak solution but it is easier, and far less dangerous, to return to the system of first-order conservation equations from which (4.1) was derived and use the ideas of Chapter 2. This eventually shows that jumps in the dependent variables of any two-by-two *linear* hyperbolic system can only propagate along the characteristics.

4.4 Transforms and eigenfunction expansions

Anyone confronted with an unfamiliar partial differential equation should always be on the lookout for any symmetry properties that may enable special methods to be used to generate explicit solutions. Indeed, we have already seen that the wave equation (4.1) is simple enough that its general solution can be written down in terms of two arbitrary functions. However, choosing these functions to satisfy required initial and/or boundary conditions may be difficult.

Fortunately, there is a powerful method for circumventing this difficulty, and it relies on the superposition principle for linear equations. The basic idea of synthesising solutions with the desired properties from elementary solutions often goes under the heading of 'Fourier representation', or 'separation of the variables', or 'transform methods', and the idea is very simple. If any linear partial differential equation, not necessarily a hyperbolic one, has a 'spectrum' of solutions u_λ whose dependence on x and y can be separated so that

$$u_\lambda = X(x,\lambda)Y(y,\lambda) \tag{4.19}$$

for some discrete or continuous parameter λ, then a summation or integral over λ may allow arbitrary initial and/or boundary conditions to be satisfied. For example, if the equation has constant coefficients, this is always a possibility because X and Y can then be exponential functions of x and y, respectively.

The tools for carrying out the summation can be quite intricate, but they are exemplified by the following two archetypal results. These both relate to the commonly occurring situation in which $X(x)$ satisfies

$$\mathcal{L}X = \frac{\mathrm{d}^2X}{\mathrm{d}x^2} = \lambda X, \tag{4.20}$$

so that λ is an eigenvalue of $\mathcal{L}$. Essentially, there are two basic possibilities.

Discrete spectrum

When we solve (4.20) on the interval $(-L, L)$ with periodic boundary data, the eigenvalues are $\lambda = -n^2\pi^2/L^2$. Then, summation of multiples of eigenfunctions X leads to the fundamental result of Fourier series, applied to suitable real functions defined on $(-L, L)$ and extended periodically. Any such function can be written as the sum of eigenfunctions of (4.20) with periodic boundary conditions in the form

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{-in\pi x/L}, \quad \text{where} \quad c_n = \frac{1}{2L}\int_{-L}^{L} f(x)e^{in\pi x/L}\,dx. \tag{4.21}$$

Continuous spectrum

When we require (4.20) to hold for all x, with X bounded as $|x| \to \infty$, the eigenvalues of (4.20) are $\lambda = -k^2$ for all real k, so that the spectrum is continuous, and we are led to the fundamental result of Fourier transforms. This result is derived formally from (4.21) by taking the limit $L \to \infty$ after writing $n\pi/L = k$ and interpreting the sum as a Riemann integral; in the limit, $2Lc_n$ is replaced by

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x)e^{ikx}\,dx \tag{4.22}$$

(called the *Fourier transform* of f), and the inverse

$$f(x) = \sum_{n=-\infty}^{\infty}\left(\frac{1}{2L}\int_{-L}^{L} f(s)e^{in\pi s/L}\,ds\right)e^{-in\pi x/L}$$

becomes

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(k)e^{-ikx}\,dk, \tag{4.23}$$

which is called the *Fourier inversion formula*. These results, summarised as

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x)e^{ikx}\,dx, \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(k)e^{-ikx}\,dk, \tag{4.24}$$

only apply to functions that behave well enough for the integrals to converge.

In more general cases, bewildering combinations of discrete and continuous spectra can occur; all the fascinating and delicate theory behind these bald statements can be found in [44].

Prompted by these results, we can summarise the key ingredients of the transform/eigenfunction expansion philosophy as follows; the next six pages contain, in condensed form, everything we will need to know about Fourier transforms and distributions, although we will have more to say about eigenfunction expansions in Chapter 5.

- *More general operators.* For general self-adjoint ordinary differential operators $\mathcal{L}$ with a discrete spectrum λ_n and real eigenfunctions $\phi_n(x)$, we expect that any real function $f(x)$ has the expansion

$$f(x) = \sum c_n\phi_n(x) \quad \text{with} \quad c_n = \int f(x)\phi_n(x)\,dx, \tag{4.25}$$

  as long as ϕ_n form a complete orthonormal set; the integral, in common with the other integrals with respect to x in this section, is over the region in which the equation is to be solved. This can be proved when the eigenvalue problem is set on a finite interval and there are no singularities in the coefficients in $\mathcal{L}$; this is *Sturm–Liouville* theory. Moreover, when the interval is infinite or semi-infinite, or when $\mathcal{L}$ has singular coefficients and has eigenvalues λ_k labelled by the continuous parameter k and corresponding orthonormal eigenfunctions $\phi(x,k)$, we need to consider the continuous transform

$$\hat{f}(k) = \int f(x)\phi(x,k)\,\mathrm{d}x, \quad f(x) = \int \hat{f}(k)\phi(x,k)\,\mathrm{d}k, \tag{4.26}$$

in which the range of the k integration is not always obvious. (Note that, when we normalise the exponential in (4.22) by multiplying by $1/\sqrt{2\pi}$, and correspondingly replace the factor $1/2\pi$ in the inversion in (4.24) by $1/\sqrt{2\pi}$, (4.21) and (4.24) are almost special cases of (4.25) and (4.26), the only discrepancy being in the sign of the arguments of the exponentials. We will resolve this shortly.) When we come to use (4.25) and (4.26) to describe the solution of any particular partial differential equation, we cannot overemphasise how important it is that the x derivatives form a self-adjoint operator $\mathcal{L}$. It is only when this is the case that multiplying the equation $\mathcal{L}f = \lambda_k f$ by $\phi(x,k)$ and integrating over x gives the inner product relation

$$(\phi, \mathcal{L}f) = (\mathcal{L}\phi, f) = \lambda_k(\phi, f), \tag{4.27}$$

so that the transform of $\mathcal{L}f$ can be written in terms of the transform of f. In the Fourier transform example above, $\lambda_k = -k^2$, each eigenvalue being double, and the orthogonal eigenfunctions are $e^{\pm ikx}$.

- *Complex eigenfunctions.* The operator $\mathcal{L} = \mathrm{d}^2/\mathrm{d}x^2$ is the simplest example of a self-adjoint operator, and this is reflected in the fact that its eigenvalues, $-n^2\pi^2/L^2$ in the discrete case and $-k^2$ in the continuous case, are real. Also, with our periodic boundary conditions, its eigenfunctions are orthogonal sine and cosine functions and the Fourier series (4.21) is often written in terms of these real functions; however, combining them into complex exponentials not only saves space but also helps conceptually. We can now see why the normalised versions of (4.21) and (4.24) referred to above do not quite agree with (4.25) and (4.26). The difference results from the fact that (4.25) and (4.26) only apply when $\phi_n(x)$ and $\phi(x,k)$ are real; when they are complex, the inner products need to be generalised to $(f,g) = \int f\overline{g}\,\mathrm{d}x$, where $\overline{g}$ is the complex conjugate of g. Also, we need to normalise so that $\int \phi_n(x)\overline{\phi}_n(x)\,\mathrm{d}x$ or $\int \phi(x,k)\overline{\phi}(x,k)\,\mathrm{d}x$ is unity, and (4.25) and (4.26) are then correspondingly replaced by

$$f(x) = \sum c_n\phi_n(x) \qquad \text{with} \qquad c_n = \int f(x)\overline{\phi}_n(x)\,\mathrm{d}x,$$
$$f(x) = \int \hat{f}(k)\phi(x,k)\,\mathrm{d}k \quad \text{with} \quad \hat{f}(k) = \int f(x)\overline{\phi}(x,k)\,\mathrm{d}x.$$

- *The Laplace transform.* Bearing in mind our encounter with causality in §4.3, it is useful to be able to represent functions that are zero for negative time (here represented by $x < 0$). In that case it is convenient to write $k = \mathrm{i}p$ and $\hat{f}(k) = \tilde{f}(p)$ to obtain, formally, the *Laplace transform* and its inverse,

$$\tilde{f}(p) = \int_0^\infty f(x)\mathrm{e}^{-px}\,\mathrm{d}x, \quad f(x) = \frac{1}{2\pi\mathrm{i}}\int \tilde{f}(p)\mathrm{e}^{px}\,\mathrm{d}p. \tag{4.28}$$

We have deliberately left the range of integration in p undefined for reasons which result from the serious convergence questions associated with the Fourier transform and its inverse. Indeed, the former does not even work for $f(x) = 1$. This is the most difficult and confusing aspect of Fourier analysis and it can be approached in the following two ways.

- *The complex Fourier transform.* The first procedure applies to the Fourier transform in its standard form (4.22) and is expounded in [8]. As in (4.28), we begin by considering functions that vanish in $x < 0$, and then we assume that they grow like $\mathrm{e}^{\alpha x}$, $\alpha > 0$, as $x \to +\infty$ (but not like $\mathrm{e}^{\alpha x^{1+\epsilon}}$ for $\epsilon > 0$). Instead of trying to find $\hat{f}(k)$, the basic idea is to consider the Fourier transform of $f(x)\mathrm{e}^{-\beta x}$, $\beta > \alpha$. The effect of introducing β is exactly the same as that of complexifying k in (4.22) and working on a contour $\Im k > \alpha$, and this involves careful contour integration (see Exercise 4.9). The principal result is that the inversion needs to be taken along a contour which lies *above* all the singularities of $\hat{f}(k)$ in the complex k plane. When the consequences are incorporated into the Laplace inversion formula, the integral needs to be taken from $\gamma - \mathrm{i}\infty$ to $\gamma + \mathrm{i}\infty$, where the real number γ is large enough for all the singularities of $\tilde{f}(p)$ to lie to its *left* in the complex p plane.
This procedure can equally be applied to functions that grow exponentially as $x \to -\infty$ and vanish in $x > 0$, allowing more general functions to be considered by extending the argument of Exercise 4.9. However, functions that grow faster than $\mathrm{e}^{\alpha x}$ as $|x| \to \infty$ can only be treated by more radical changes in the contour integrals in (4.22). When faced with, say, $f(x) = \mathrm{e}^{x^2}$, we define its transform as

$$F(\zeta) = \int_\Gamma \mathrm{e}^{\zeta x}\mathrm{e}^{x^2}\,\mathrm{d}x,$$

where Γ is any contour starting at infinity in $5\pi/4 < \arg x < 7\pi/4$ and ending at infinity in $\pi/4 < \arg x < 3\pi/4$. We soon find that $F(\zeta) = \mathrm{i}\sqrt{\pi}\mathrm{e}^{-\zeta^2/4}$, and the inversion formula is

$$\frac{1}{2\pi\mathrm{i}}\int_{\tilde{\Gamma}} \mathrm{i}\sqrt{\pi}\mathrm{e}^{-\zeta^2/4}\mathrm{e}^{-\zeta x}\,\mathrm{d}\zeta,$$

where $\tilde{\Gamma}$ is again chosen so that the integral exists. Indeed, the representation

$$f(x) = \frac{1}{2\pi\mathrm{i}}\int_{\tilde{\Gamma}} \mathrm{e}^{-\zeta x}F(\zeta)\,\mathrm{d}\zeta$$

is a useful general technique for solving certain ordinary differential equations (see Exercise 4.10).

- *Fourier transforms and distributions.* The second procedure is simply to permit series such as (4.21) and integrals such as (4.22) to diverge in the usual sense but to interpret them as generalised functions or distributions, as discussed on p. 98. For example, the periodically extended function

$$f(x) = \begin{cases} -1, & -L < x < 0, \\ 1, & 0 < x < L, \end{cases}$$

has the Fourier series

$$f(x) = \sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi} \sin \frac{(2n+1)\pi x}{L}.$$

Assuming that $f'(x) = 2\delta(x)$ in $-L < x < L$ (the delta function comes from the derivative of the jump at $x = 0$), it is tempting to write

$$f'(x) = \sum_{n=0}^{\infty} \frac{4}{L} \cos \frac{(2n+1)\pi x}{L} = 2 \sum_{n=-\infty}^{\infty} (-1)^n \delta(x - nL).$$

Fortunately, this outrageous-looking statement is easy to justify. A series $\sum D_n$ of distributions is said to converge if the series of numbers $\sum (D_n, \psi)$ converges in the usual sense whenever ψ is a test function. Term-by-term differentiation is justified because $\sum (D'_n, \psi) = -\sum (D_n, \psi')$, and if ψ is a test function then so is ψ', so the right-hand side converges. All the difficulties associated with the 'usual' convergence of series disappear here, because the test functions are assumed to be so smooth.

This idea of defining properties of generalised functions by transferring them to the test functions can be carried further to define the Fourier transform of a distribution. The key here is the formal observation that, for functions $f(x)$ and $g(x)$,

$$\begin{aligned} (\hat{f}, g) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x) \mathrm{e}^{ikx}\, \mathrm{d}x \right) g(k)\, \mathrm{d}k \\ &= \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} g(k) \mathrm{e}^{ikx}\, \mathrm{d}k \right) \mathrm{d}x \\ &= (f, \hat{g}), \end{aligned}$$

as long as the orders of integration can be changed (this is called *Parseval's formula*). This suggests that we define the Fourier transform of a distribution D to be $\hat{D}(k)$, where

$$(\hat{D}, \psi) = (D, \hat{\psi}),$$

with the corresponding formula

$$(\check{D}, \psi) = (D, \check{\psi})$$

for the inverse $\check{D}(x)$, where $\check{\psi}$ is the inverse Fourier transform of ψ as given by (4.24). When this is done, a beautiful theory emerges in which most useful distributions have suitably defined Fourier transforms that are themselves distributions. This theory enables us easily to confirm results such as

$$\hat{\delta}(k) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}kx}\delta(x)\,\mathrm{d}x = 1, \quad \hat{1} = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x = 2\pi\delta(k),$$

and hence that the inversion formula for $\hat{f}(k)\hat{g}(k)$ is

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(k)\hat{g}(k)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}k \\
&\quad = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(y)\mathrm{e}^{\mathrm{i}ky}\,\mathrm{d}y\right)\left(\int_{-\infty}^{\infty} g(z)\mathrm{e}^{\mathrm{i}kz}\,\mathrm{d}z\right)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}k \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)g(z)\delta(x-y-z)\,\mathrm{d}y\,\mathrm{d}z \\
&\quad = \int_{-\infty}^{\infty} f(y)g(x-y)\,\mathrm{d}y, \qquad (4.29)
\end{aligned}$$

the so-called *convolution formula*; the last integral is often written as $f * g$.

- *Multidimensional Fourier transforms.* We can formally generalise (4.22) to functions $f(x,y)$ by defining

$$\hat{f}(k_1,k_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{e}^{\mathrm{i}(k_1x+k_2y)}\,\mathrm{d}x\,\mathrm{d}y, \qquad (4.30)$$

so that (4.23) suggests that

$$f(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{f}(k_1,k_2)\mathrm{e}^{-\mathrm{i}(k_1x+k_2y)}\,\mathrm{d}k_1\,\mathrm{d}k_2, \qquad (4.31)$$

and the extension to more variables is obvious. Cases where the inversion can be done explicitly are even rarer than for functions of one variable, and the question of convergence becomes even more complicated. Nevertheless, we will see that quite simple changes of variable in the integrals in (4.30) or (4.31) can sometimes give us valuable insights. A spectacular example occurs when we consider $\hat{f}$ as a function of k and θ, where $k_1 = k\cos\theta$ and $k_2 = k\sin\theta$, so that $\hat{f}$ is equal to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{e}^{\mathrm{i}k(x\cos\theta+y\sin\theta)}\,\mathrm{d}x\,\mathrm{d}y.$$

When we rotate the axes by writing

$$x = r\cos\theta - t\sin\theta, \quad y = r\sin\theta + t\cos\theta,$$

this becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ikr} f(r\cos\theta - t\sin\theta, r\sin\theta + t\cos\theta)\, dr\, dt.$$

Hence, as a function of k and θ, $\hat{f}$ is equal to

$$\int_{-\infty}^{\infty} e^{ikr} f_R(r,\theta)\, dr,$$

where

$$f_R(r,\theta) = \int_{-\infty}^{\infty} f(r\cos\theta - t\sin\theta, r\sin\theta + t\cos\theta)\, dt. \tag{4.32}$$

This is called the *Radon transform* of $f(x,y)$ and it is fundamental to the analysis of CAT scanning tomography. Suppose we shine a thin beam of X-rays through a two-dimensional body at an angle θ; the transmitted intensity is then determined by the integral of the absorption coefficient $f(x,y)$ along the beam, that is by the Radon transform of f in this direction. Repeating the scan at the same angle θ along parallel beams for each value of the lateral coordinate r, and then scanning again for all $0 \leqslant \theta < \pi$, we find $f_R(r,\theta)$. Then, in order to retrieve $f(x,y)$, all we need to do is to take the Fourier transform in r of $f_R(r,\theta)$, transform into Cartesian coordinates, and invert $\hat{f}$.
We must emphasise one severe limitation on the use of distributions in multiple Fourier transforms. This is that, while we can perfectly easily define the Fourier transform of $\delta(x)\delta(y)$ as a product of two one-dimensional transforms (which is hence equal to unity), there is no satisfactory definition of, say, $\delta(x)\delta(x)$. This remark applies to any product of distributions with the same 'independent variable' unless, of course, they are functions in the usual sense.

- *The transform integral regarded as an integral equation.* We might consider another approach to the problem of recovering a function from its transform: instead of using the inversion formula, we might try to find $f(x)$, given its transform $\hat{f}(k)$, by solving

$$\int f(x)\phi(x,k)\, dx = \hat{f}(k)$$

as an integral equation for $f(x)$. Now readers familiar with the theory of integral equations will be aware that Fredholm equations of the first kind, in which a function $f(x)$ has to be determined such that

$$\mathcal{K}[f](t) = \int_a^b K(x,t)f(x)\, dx = g(t), \tag{4.33}$$

for given $K(x,t)$ and $g(t)$, are usually ill-posed. One simple-minded reason for this is that, if K is separable, so that $K(x,t) = \sum_1^N \alpha_n(x)\beta_n(t)$, where $N < \infty$, then (4.33) can never have a solution unless $g(t)$ is a linear combination of $\beta_n(t)$; moreover, if any functions $f_e(x)$ exist such that $\int_a^b K(x,t)f_e(x)\, dx = 0$, then, even if $f(x)$ exists, it is indeterminate to within a linear combination of these functions. More generally, if K is not separable and $\int_a^b\int_a^b |K(x,t)|^2\, dx\, dt < \infty$, it is a standard result that, as long as a and b are finite, there exists an infinite sequence of eigenvalues λ_n and a complete sequence of eigenfunctions $f_n(x)$ such that

$$\lambda_n \int_a^b K(x,t) f_n(x)\,\mathrm{d}x = f_n(t),$$

and, moreover, $\lambda_n \to \infty$ as $n \to \infty$. Because these eigenfunctions are complete, every well-behaved function can be expanded as a generalised Fourier series in terms of f_n. Now suppose that $f(x) = \sum a_n f_n(x)$; then the operator $\mathcal{K}$ is such that

$$\mathcal{K}[f](t) = \sum \frac{a_n}{\lambda_n} f_n(t),$$

which is a series whose coefficients are ultimately smaller in modulus than the original ones and thus represents a 'smaller' function than the input $f(x)$ (for this reason such operators are called *compact*). Hence, if we try to solve (4.33) for a compact operator $\mathcal{K}$ by expanding $g(t) = \sum b_n f_n(t)$ and equating coefficients, we find $a_n = \lambda_n b_n$. That is, f is *less* smooth than g, and so we can only hope to find a solution if we restrict the class of possible right-hand sides $g(t)$; this is another indication of ill-posedness.

These observations might cause us to doubt the value of (4.24); the last thing we want is for Fourier inversion to be ill-posed! However, we are reassured by the facts that the kernel $\phi(x,k) = \mathrm{e}^{ikx}$ is not separable, the corresponding operator $\mathcal{K}$ has no null space and, most importantly, $\mathcal{K}$ is not compact (to see this, consider the 'small' function $f(x) = \epsilon\mathrm{e}^{-\epsilon|x|}$, which converges uniformly to zero as $\epsilon \to 0$; its Fourier transform is $\hat{f}(k) = 2\epsilon^2/(\epsilon^2 + k^2)$ which, however, does not tend uniformly to zero).

4.5 Applications to wave equations

From the practical point of view, the message that emerges from the discussion of the previous section is that linear partial differential equations (hyperbolic or otherwise) that are 'symmetric' enough to have a spectrum of solutions (4.19) can be solved by generalised Fourier methods, that the particular method will depend on the problem in hand, and that technical pitfalls may well be encountered concerning questions of convergence. With this in mind let us now consider some simple hyperbolic problems.

4.5.1 The wave equation in one space dimension

Let us consider

$$\frac{\partial^2 u}{\partial t^2} = a_0^2 \frac{\partial^2 u}{\partial x^2} \quad \text{for } t > 0. \tag{4.34}$$

We have already given the d'Alembert solution to the Cauchy problem with data at $t = 0$ in (3.31) and Exercise 3.8, but the constancy of the coefficients in (4.34) also makes it suitable for Fourier analysis. If, for example,

$$u = 0, \quad \frac{\partial u}{\partial t} = v_0(x) \quad \text{at } t = 0,\ -\infty < x < \infty,$$

the range of the respective independent variables indicates that the problem is amenable to a Fourier transform in x or a Laplace transform in t (or both). The result of the former is

$$\frac{d^2\hat{u}}{dt^2} + a_0^2k^2\hat{u} = 0,$$

and so

$$\hat{u}(k,t) = \hat{f}(k)e^{ika_0t} + \hat{g}(k)e^{-ika_0t} \tag{4.35}$$

for some $\hat{f}$ and $\hat{g}$; incorporating the initial data we find

$$\hat{u}(k,t) = \frac{\hat{v}_0(k)}{a_0k}\sin a_0kt, \tag{4.36}$$

from which

$$u(x,t) = \frac{1}{2\pi a_0}\int_{-\infty}^{\infty}\frac{\hat{v}_0(k)}{2ik}\left(e^{ik(a_0t-x)} - e^{-ik(a_0t+x)}\right)dk, \tag{4.37}$$

with suitable assumptions about the behaviour of v_0 and $\hat{v}_0$. We can write this as

$$u(x,t) = \frac{1}{2\pi a_0}\int_{-\infty}^{\infty} v_0(s)\left(\int_{-\infty}^{\infty}\frac{e^{ik(s+a_0t-x)} - e^{ik(s-a_0t-x)}}{2ik}\,dk\right)ds$$

and use the result from contour integration that, when ν is real,

$$\int_{-\infty}^{\infty}\frac{e^{i\nu k}}{k}\,dk = \pm i\pi$$

for $\nu > 0$ and $\nu < 0$, respectively (no matter which way we indent the contour at $k = 0$), to show that the integral with respect to k vanishes, unless $x - a_0t < s < x + a_0t$ when it is π. Hence (4.37) is in accordance with the d'Alembert solution

$$u(x,t) = \frac{1}{2}\left(u_0(x-a_0t) + u_0(x+a_0t)\right) + \frac{1}{2a_0}\int_{x-a_0t}^{x+a_0t} v_0(s)\,ds, \tag{4.38}$$

which applies to this problem when $u(x,0) = u_0$ as well as $\partial u/\partial t(x,0) = v_0$. Thus, we have effectively retrieved (3.31), albeit by a more long-winded but more powerful method. However, (4.37) illustrates how concepts such as regions of influence and domains of dependence can emerge naturally from Fourier representations; outside a region of influence the solution is represented by an integral round a contour containing no interior singularities, and is thus zero.

We also remark that (4.37) is a very typical result in the application of Fourier transforms. In the case of this Cauchy problem, the method effectively retrieves the general solution of the wave equation in terms of the arbitrary functions u_0 and v_0. Of course, Fourier representations such as (4.37) satisfy the wave equation for *arbitrary* initial and boundary conditions, but often it is not so easy to relate the arbitrary functions of k that enter into the Fourier representation to the initial/boundary data for any particular problem.

The alternative Laplace transform approach is even more messy: we obtain

$$a_0^2 \frac{\mathrm{d}^2 \tilde{u}}{\mathrm{d}x^2} - p^2 \tilde{u} = -v_0(x),$$

for which the appropriate solution is

$$\tilde{u}_0(x,p) = \frac{1}{2a_0 p}\left(\int_{-\infty}^{x} \mathrm{e}^{p(\xi-x)/a_0} v_0(\xi)\,\mathrm{d}\xi + \int_{x}^{\infty} \mathrm{e}^{-p(\xi-x)/a_0} v_0(\xi)\,\mathrm{d}\xi\right), \quad \Re p > 0.$$

A formal reverse of the order of integration in the Laplace inversion gives the d'Alembert solution (4.38) with $u_0 = 0$ when we remember that

$$\frac{1}{2\pi \mathrm{i}} \int_{\gamma - \mathrm{i}\infty}^{\gamma + \mathrm{i}\infty} \mathrm{e}^{p(\xi-x)/a_0} \mathrm{e}^{pt} \frac{\mathrm{d}p}{p} = \begin{cases} 0, & x - \xi > a_0 t, \\ 1, & x - \xi < a_0 t, \end{cases}$$

when $\Re\gamma > 0$.

The initial/boundary value problem in which

$$u = 0, \quad \frac{\partial u}{\partial t} = v_0(x), \quad \text{for } 0 < x < 1, \tag{4.39}$$

with, say, $u = 0$ at $x = 0$ and $x = 1$, which is the two-boundary version of the problem we presented in Fig. 4.3, can be attempted by either a Fourier series in x (because $X(x, \lambda) = \sin\left(\sqrt{-\lambda}\pi x\right))$ or a Laplace transform in t. Using the specialisation of (4.21) to odd functions, so as to satisfy the boundary conditions,[43] we find

$$u = \sum_{n=1}^{\infty} b_n \sin n\pi x \sin n\pi a_0 t, \tag{4.40}$$

where

$$b_n = \frac{2}{n\pi a_0} \int_0^1 v_0(x) \sin n\pi x \,\mathrm{d}x,$$

which is sometimes a more convenient description of the solution than would be the equivalent of (4.38), encompassing as it does all the possible discontinuities running along the 'zig-zag' characteristics in Fig. 4.5. The bottom left-hand corner of Fig. 4.5 is Fig. 4.3.

We also remark that we can find the general solution of (4.34) by superposing elementary solutions of the form $\delta(x - a_0 t)$ and $\delta(x + a_0 t)$. Instead of (4.35), we obtain

$$\begin{aligned} u(x,t) &= \int_{-\infty}^{\infty} (\delta(x - x' - a_0 t) f(x') + \delta(x - x' + a_0 t) g(x'))\,\mathrm{d}x' \\ &= f(x - a_0 t) + g(x + a_0 t), \end{aligned} \tag{4.41}$$

where f and g are arbitrary; another route to this formula is given in Exercise 4.8, and we will return to this observation below.

[43]This can be systematised by saying that we seek $\mathcal{L}$ and X as in (4.20), with $X(0) = X(1) = 0$.

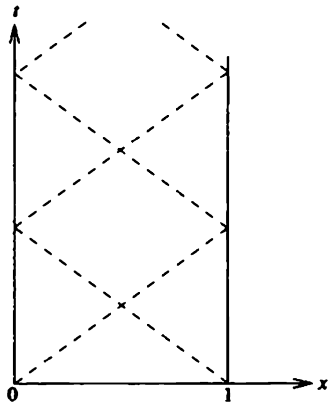


Fig. 4.5 Two-point boundary value problem for the wave equation.

4.5.2 Circular and spherical symmetry

We will deal with multidimensional wave propagation more seriously in the next section, but when we have circular symmetry in two space dimensions the wave equation is

$$\frac{\partial^2 u}{\partial t^2} = a_0^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r}\right). \tag{4.42}$$

If we generalise the shallow water model of §2.1 and linearise about a state of rest as in §4.1, we can show that this models the waves that are generated on the surface of a shallow pond when, for example, a stone is thrown in. For the Cauchy problem

$$u = u_0(r), \quad \frac{\partial u}{\partial t} = v_0(r) \quad \text{at } t = 0, 0 < r < \infty, \tag{4.43}$$

we can still use a Laplace transform in t. However, the relevant eigenvalue problem in spatial variables when u is time harmonic with frequency ω, i.e. $u = \Re\left(R(r)\mathrm{e}^{-\mathrm{i}\omega t}\right)$, is

$$\mathcal{L}R = \frac{\mathrm{d}^2 R}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}R}{\mathrm{d}r} = \lambda R, \tag{4.44}$$

in which $\lambda = -\omega^2/a_0^2$ and R is finite at $r = 0$ and as $r \to \infty$, and it is not self-adjoint. In the light of the argument given after (4.25) we must first arrange that the spatial eigenvalue problem is self-adjoint by rewriting (4.42) as

$$\frac{\partial^2 (ru)}{\partial t^2} = a_0^2\left(r\frac{\partial^2 u}{\partial r^2} + \frac{\partial u}{\partial r}\right) = a_0^2\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right). \tag{4.45}$$

Then, (4.44) becomes

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}R}{\mathrm{d}r}\right) = \lambda r R,$$

which has well-behaved solutions at $r = 0$ when λ is real and negative. We write $\lambda = -k^2$, so that R is proportional to $J_0(kr)$, the Bessel function of first kind and zeroth order. From (4.26) the appropriate spatial transformation is the so-called *Hankel transform*

$$\hat{u}(k,t) = \int_0^\infty r u(r,t) J_0(kr)\,\mathrm{d}r.$$

The inversion formula for this transform is found most easily by taking the double Fourier transform of a function $U(x,y,t)$ with respect to x and y, namely

$$\hat{U}(k_1,k_2,t) = \int_{-\infty}^\infty \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}(k_1x+k_2y)} U(x,y,t)\,\mathrm{d}x\,\mathrm{d}y, \tag{4.46}$$

for which (4.31) gives

$$U(x,y,t) = \frac{1}{4\pi^2}\int_{-\infty}^\infty \int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}(k_1x+k_2y)} \hat{U}(k_1,k_2,t)\,\mathrm{d}k_1\,\mathrm{d}k_2.$$

When $U(x,y,t) = u(r,t)$, (4.46) becomes

$$\begin{aligned}\hat{U}(k_1,k_2,t) &= \int_0^\infty r u(r,t) \int_0^{2\pi} \mathrm{e}^{\mathrm{i}kr\cos\theta}\,\mathrm{d}\theta\,\mathrm{d}r, \quad \text{where} \quad k^2 = k_1^2 + k_2^2,\\ &= 2\pi\hat{u}(k,t),\end{aligned}$$

since, as can be verified by direct differentiation,

$$\frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}kr\cos\theta}\,\mathrm{d}\theta = J_0(kr). \tag{4.47}$$

Similarly,

$$u(r,t) = \frac{1}{2\pi}\int_0^\infty k\hat{u}(k,t)\left(\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}kr\cos\phi}\,\mathrm{d}\phi\right)\mathrm{d}k$$

and so[44]

$$u(r,t) = \int_0^\infty k\hat{u}(k,t) J_0(kr)\,\mathrm{d}k.$$

If, for simplicity, we take $u_0 = 0$, we find that

$$u(r,t) = \frac{1}{a_0}\int_0^\infty \hat{v}_0(k) J_0(kr) \sin a_0kt\,\mathrm{d}k; \tag{4.48}$$

this formula has important repercussions to which we will return later.

In three space dimensions, (4.42) is replaced by

$$\frac{\partial^2 u}{\partial t^2} = a_0^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\right), \tag{4.49}$$

and now a surprising 'symmetry' occurs. By writing $u = rv$ we find

$$\frac{\partial^2 v}{\partial t^2} = a_0^2\frac{\partial^2 v}{\partial r^2},$$

and we are back in one space dimension. Thus, for hyperbolic problems, it is sometimes easier to proceed when there are more independent variables than when there are fewer, as we shall also discuss further in the next section.

[44]We will derive this 'transform pair' by another method in Chapter 5.

*4.5.3 The telegraph equation

By the use of the exponential transformation mentioned at the end of §4.2.3, the telegraph equation can always be considered in the form

$$a_0^2\frac{\partial^2 u}{\partial x^2}-\frac{\partial^2 u}{\partial t^2}-cu=0, \tag{4.50}$$

where $c>0$ in practical problems, and the Cauchy problem with u and $\partial u/\partial t$ prescribed for all x at $t=0$ is suitable for a Fourier transform. When u is zero initially, the analogue of (4.36) is

$$\hat{u}=\frac{\hat{v}_0(k)}{\sqrt{a_0^2k^2+c}}\sin\left(\sqrt{a_0^2k^2+c}\right)t, \tag{4.51}$$

the apparent branch points at $k=\pm\mathrm{i}\sqrt{c}/a_0$ being illusory (removable). The details of the identification of the Fourier inversion of (4.51) with the Riemann function representation of the solution derived in §4.2.3 are given in Exercise 4.12, but we can give a quicker way of removing the guesswork used in deriving (4.15). Taking $\xi=\eta=0$, we seek a Riemann function directly, satisfying

$$\frac{\partial^2 R}{\partial x\partial y}+R=\delta(x)\delta(y)$$

with R non-zero only in $x\leqslant 0$, $y\leqslant 0$. Then the Fourier transform of R with respect to x satisfies

$$\frac{\partial\hat{R}}{\partial y}+\frac{\mathrm{i}\hat{R}}{k}=\frac{\mathrm{i}}{k}\delta(y)$$

and so

$$\hat{R}=-\frac{\mathrm{i}}{k}\mathrm{e}^{-\mathrm{i}y/k}H(-y),$$

where H is the Heaviside function. Hence

$$R(x,y)=-\frac{\mathrm{i}H(-y)}{2\pi}\int_{\Im k=\text{constant}<0}\mathrm{e}^{-ikx-\mathrm{i}y/k}\frac{\mathrm{d}k}{k},$$

the inversion contour being chosen as described in §4.4. This integral clearly vanishes when $x>0$, by closing the contour in $\Im k<0$, but if $x<0$ we can write it as

$$R(x,y)=-\frac{\mathrm{i}H(-y)}{2\pi}\int_{\Im k'=\text{constant}<0}\mathrm{e}^{-\mathrm{i}\sqrt{xy}(k'+1/k')}\frac{\mathrm{d}k'}{k'}.$$

Now we deform the contour into the unit circle around the origin so that, when x and y are both negative,

$$R(x,y)=\frac{1}{2\pi}\int_0^{2\pi}\mathrm{e}^{-2\mathrm{i}\sqrt{xy}\cos\theta}\,\mathrm{d}\theta=J_0\left(2\sqrt{xy}\right).$$

It is interesting at this stage to recall our discussion at the end of §4.2.2 about the propagation of discontinuities in the Riemann function. For the wave equation,

$R \equiv 1$ in $x \leqslant 0$, $y \leqslant 0$ which, by differentiation, means that solutions are possible in which u is precisely a sum of delta functions on $x = 0$ and $y = 0$. For the telegraph equation, however, such delta functions can only propagate if there is a non-zero wave field between the characteristics $x = 0$ and $y = 0$. We will discuss this point further in the next section and in Exercise 4.12.

*4.5.4 Waves in periodic media

The study of the solutions of

$$\frac{\partial^2 u}{\partial t^2} = a_0^2(x)\frac{\partial^2 u}{\partial x^2}, \tag{4.52}$$

where a_0 is periodic in x, is important for many practical problems, for example electromagnetic wave propagation in crystalline solids such as semiconductors. The relevant spatial eigenvalue problem, when u is harmonic in time with frequency ω, is

$$\frac{\mathrm{d}^2 X}{\mathrm{d}x^2} + \frac{\omega^2}{a_0^2(x)}X = 0, \tag{4.53}$$

with ω real, so there are few explicit solutions for such problems. Nonetheless, (4.53) is susceptible to ***Floquet theory*** [2], which roughly speaking says that as ω increases the qualitative behaviour of X alternates; when ω lies in some intervals, called ***stop bands***, the two independent solutions of (4.53) either grow or decay as $|x| \to \infty$, but when ω lies in the complementary 'pass band' intervals these eigensolutions are all quasiperiodic in x, and are hence physically acceptable for problems in which (4.52) holds in $-\infty < x < \infty$. It is only at the boundaries between the pass and stop bands that X can be periodic. Hence, even without an explicit solution in front of us, we can see a new phenomenon that can occur for hyperbolic equations: waves may be able to propagate along characteristics but their spatial penetration may be much greater when frequencies lie in certain bands than in others. In fact, such ***dispersive*** behaviour is quite general, because, even for the telegraph equation (4.50), which has constant coefficients, waves with frequency ω are such that the wavenumber

$$k = \sqrt{\frac{\omega^2}{a_0^2} - c},$$

and hence the solution decays or grows exponentially when $|\omega| < a_0\sqrt{c}$.

*4.5.5 General remarks

Continuing to catalogue examples like this would entrain us ever deeper into problem-specific technicalities. However, all the examples we could have displayed would reveal conformity with the predictions of the Riemann function representation (4.14). They all have domains of dependence and regions of influence, the only example that might have given us pause for thought in this respect being the telegraph equation (4.50). This equation does admit solutions of the form

$$u = \Re\, \mathrm{e}^{ik(x-Vt)},$$

where

$$V = \sqrt{a_0^2 + \frac{c}{k^2}}, \tag{4.54}$$

and, in the physically realisable case $c > 0$, this appears to predict the propagation of waves travelling faster than the characteristic speed a_0. However, what happens in any practical initial value problem where, say, Cauchy data at $t = 0$ has compact support (i.e. vanishes outside a finite range of x) is that the Fourier components with wavenumber k can disperse, i.e. rearrange themselves *within* the region of influence of the Cauchy data at a speed greater than a_0; the representation (4.14) shows that none of this rearrangement is felt outside this region of influence. Another way of looking at this is to say that the discontinuities that inevitably occur on the boundary of the region of influence are described by very large values of k and, from (4.54), this boundary has speed a_0.

This raises the general question of the way in which the solution sorts itself out within the region of influence of the Cauchy data. We can see from Exercise 4.22 that, for the one-dimensional wave equation (4.1) (and hence the three-dimensional radially-symmetric equation (4.49)), any initially localised data for u with $\partial u/\partial t\,(x, 0) = 0$ gives a response that varies just in the vicinity of the characteristics $x = \pm a_0 t$ emanating from the source. Our region of influence argument led us to anticipate that there is no disturbance in $x > a_0 t$ or $x < -a_0 t$, but it did not reveal that in between these characteristics there is a 'zone of silence' in which u is zero (in three dimensions, u vanishes in $0 \leqslant r < a_0 t$). However, when we consider the circularly-symmetric wave equation (4.42) or the telegraph equation (4.50) with localised Cauchy data, we can show that the solution is non-zero *(almost) everywhere* in the region $0 \leqslant r < a_0 t$ (and this is borne out by the motion of a leaf floating on a pond when a stone is thrown nearby). We have already hinted at this at the end of §4.5.3, and we will have more to say about it in the next section.

4.6 Wave equations with more than two independent variables

Turning to wave equations with three or more independent variables, there are two related ideas that we have already touched upon but wish to discuss in greater detail in this section before we frame our notion of hyperbolicity in this more general case.

4.6.1 The method of descent and Huygens' principle

The ideas of this section are best illustrated by the now-familiar acoustic model

$$\frac{\partial^2 u}{\partial t^2} = a_0^2 \nabla^2 u \tag{4.55}$$

in three space dimensions, for which we have shown via (4.49) that the general radially-symmetric incoming and outgoing solutions are

$$u = \frac{1}{r} F(r \pm a_0 t), \tag{4.56}$$

respectively, when r is a spherical polar coordinate. These solutions all decrease as r increases with $r \pm a_0 t$ fixed, which reflects the merciful spreading out of sound waves in three-dimensional space; they are all also singular at $r = 0$, representing the possibility of focusing, which is absent in one-dimensional problems. Alas, no representation as simple as (4.56) exists in two space dimensions, for reasons which will soon become apparent. However, we can use (4.56) to generate the explicit solution of the general Cauchy problem[45] in which

$$u = u_0(\mathbf{x}), \quad \frac{\partial u}{\partial t} = v_0(\mathbf{x}) \quad \text{at } t = 0, \tag{4.57}$$

where $\mathbf{x} = (x, y, z)$. As in (4.41), we proceed by superimposing weighted solutions of the form (4.56) in the case where F is localised and, by causality, we only take outgoing waves; in other words we write

$$u(x,y,z,t) = \iiint \frac{\delta(r - a_0 t)}{r} f(x', y', z')\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z', \tag{4.58}$$

where $r^2 = (x-x')^2 + (y-y')^2 + (z-z')^2$, which clearly satisfies (4.55) for arbitrary weight functions f. The presence of the delta function in the integral ensures that we need only account for values of f on the surface of a sphere with the centre (x, y, z) and radius $a_0 t$. Hence, changing to spherical polar coordinates (r, θ, ϕ) and integrating with respect to r, we obtain a surface integral over the sphere of radius $a_0 t$. Remembering that the surface element is $a_0^2 t^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$, we obtain

$$u = 4\pi a_0 t\,\mathcal{L}f,$$

where

$$\mathcal{L}f = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(x + a_0 t \sin\theta\cos\phi, y + a_0 t \sin\theta\sin\phi, z + a_0 t\cos\theta) \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi. \tag{4.59}$$

For this ***retarded potential*** solution, it is clear that

$$u = 0 \quad \text{at } t = 0 \qquad \text{and} \qquad \left.\frac{\partial u}{\partial t}\right|_{t=0} = 4\pi a_0\,\mathcal{L}f|_{t=0} = 4\pi a_0 f(x, y, z),$$

by direct integration. Hence, setting $f = v_0/(4\pi a_0)$ gives us the solution to (4.57) when $u_0 = 0$. However, we can solve the complete Cauchy problem by making the crafty observation that if u satisfies (4.55) then so does $\partial u/\partial t$. Thus, $\partial/\partial t\,(t\mathcal{L}f)$ is a solution of (4.55) for all $t > 0$ and, by the argument above, its value at $t = 0$ is $4\pi a_0 f$; simply taking $f = (1/4\pi a_0)u_0$ enables the first part of (4.57) to be satisfied. Finally, we note that $\partial^2/\partial t^2\,(t\,\mathcal{L}f) = a_0^2 t \nabla^2 \mathcal{L}f$, since $t\,\mathcal{L}f$ satisfies (4.55), and hence vanishes at $t = 0$. Thus, we have solved the Cauchy problem in the form

$$u = t\,\mathcal{L}v_0 + \frac{\partial}{\partial t}(t\,\mathcal{L}u_0).$$

[45]This problem can be shown to have a unique solution by an energy argument similar to that in Exercise 4.13.

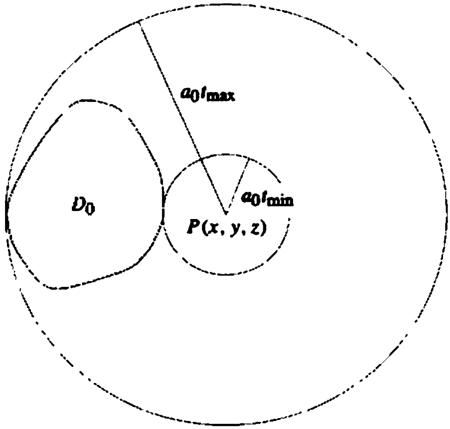


Fig. 4.6 Huygens' principle.

In interpreting this expression for the solution, we note that, in the integral (4.59), the only values of v_0 that appear are on the surface of a sphere $S_P(t)$ of radius a_0t and centre (x, y, z) (denoted by P in Fig. 4.6). Hence, if both u_0 and v_0 are only non-zero in a finite domain D_0 and P is not in D_0, then the solution obtained is zero until t is large enough, say $t > t_{\min}$, for $S_P(t)$ to intersect D_0. Also, the solution is again zero for values of t large enough, say $t \geqslant t_{\max}$, that $S_P(t)$ no longer intersects D_0, so that D_0 is totally inside $S_P(t)$. The solution is therefore non-zero in D_0 only for $t_{\min} < t < t_{\max}$; the disturbance prescribed in D_0 at $t = 0$ propagates with a *distinct* leading and trailing wave-front. This is a manifestation of what is called *Huygens' principle*.

As in §4.5.1, it is interesting to compare this procedure with that of taking a Fourier transform, again for the special case in which

$$u(\mathbf{x}, 0) = 0, \quad \frac{\partial u}{\partial t}(\mathbf{x}, 0) = v_0(\mathbf{x}).$$

We can see at once that the Fourier transform of $u(x, y, z, t)$ is

$$\hat{u}(k_1, k_2, k_3, t) = \hat{v}_0(k_1, k_2, k_3)\frac{\sin a_0kt}{a_0k},$$

where $k^2 = k_1^2 + k_2^2 + k_3^2$. Hence, whereas (4.58) is the three-dimensional generalisation of (4.41), the corresponding analogue of (4.37) is

$$u(x, y, z, t) = \frac{1}{8\pi^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(k_1x+k_2y+k_3z)}\hat{v}_0(k_1, k_2, k_3)\frac{\sin a_0kt}{a_0k}\,dk_1\,dk_2\,dk_3. \tag{4.60}$$

Moreover, we can retrieve (4.58) by substituting for $\hat{v}_0$ in terms of v and writing

$$u(x,y,z,t) =$$
$$\frac{1}{8\pi^3 a_0}\iiint v_0(x',y',z')\left(\iiint \sin a_0 kt\, \mathrm{e}^{-ikr\cos\theta} k\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}k\right)\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z',$$

where $r^2 = (x-x')^2+(y-y')^2+(z-z')^2$ and (k,θ,ϕ) are spherical polar coordinates centred at $k_1 = k_2 = k_3 = 0$. Hence, integrating with respect to θ and ϕ,

$$\begin{aligned} u(x,y,z,t) &= \frac{1}{4\pi^2 a_0}\iiint v_0(x',y',z')\left(\int_{-\infty}^{\infty}\frac{\sin a_0 kt\sin kr}{r}\,\mathrm{d}k\right)\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z' \\ &= \frac{1}{4\pi a_0}\iiint v_0(x',y',z')\frac{\delta(r-a_0t)}{r}\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'. \end{aligned}$$

To obtain the solution for a two-dimensional initial value problem we may use the *method of descent* on the result (4.59). If we assume that u_0 and v_0 are independent of z and D_0 is an infinite cylinder with generators in the z direction, then

$$\mathcal{L}v_0 = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} v_0(x+a_0t\sin\theta\cos\phi,\ y+a_0t\sin\theta\sin\phi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

With the substitution $\rho = a_0t\sin\theta$, so that $|\rho| < a_0t$,

$$\mathcal{L}v_0 = \frac{1}{2\pi a_0 t}\int_0^{2\pi}\int_0^{a_0t} v_0(x+\rho\cos\phi,\ y+\rho\sin\phi)\frac{\rho\,\mathrm{d}\rho\,\mathrm{d}\phi}{(a_0^2t^2-\rho^2)^{1/2}},$$

and ρ and ϕ are two-dimensional polar coordinates. Hence, in Cartesian coordinates $\xi = \rho\cos\phi + x$ and $\eta = \rho\sin\phi + y$, this reduces to

$$\mathcal{L}v_0 = \frac{1}{2\pi a_0 t}\iint \frac{v_0(\xi,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta}{(a_0^2t^2-(x-\xi)^2-(y-\eta)^2)^{1/2}}, \tag{4.61}$$

where the integration is now over $|\rho| < a_0t$. If the initial data is defined and nonzero on a region A_0 (the two-dimensional cross-section of D_0) then, for all $t > t_{\min}$, the interior of the circle of radius a_0t intersects A_0, and the integral (4.61) is nonzero. This ties in with our remark at the end of §4.5.5 about leaves floating on the surface of a shallow pond, and it is consistent with the prolonged rumbling heard after the initial clap of thunder caused by lightning strikes which, although jagged, are sufficiently elongated to generate an approximately two-dimensional sound field.

A further application of the method of descent reduces (4.61) to the d'Alembert formula and both leading and trailing wave-fronts occur in one space dimension for initial data with compact support, i.e. vanishing outside a finite interval.

We remark that the ideas above can also be used on inhomogeneous wave equations of the form

$$\frac{\partial^2 u}{\partial t^2} - a_0^2\nabla^2 u = f(\mathbf{x},t). \tag{4.62}$$

As observed by Duhamel, if we let $v(\mathbf{x},t,\tau)$ be the solution of the homogeneous equation ($f\equiv 0$) with $v = 0$ and $\partial v/\partial t = f(\mathbf{x},\tau)$ at $t=\tau$, solve for v using (4.59), and set $u = \int_0^t v(\mathbf{x}, t, \tau)\,\mathrm{d}\tau$, we satisfy (4.62) with zero Cauchy data. This simply reflects the fact that the source term f is equivalent to a superposition of 'pulsed' initial value problems.

It is interesting to compare the ease with which (4.61) can be obtained by the method of descent with the difficulty that confronts a Fourier analysis. Suppose, for example, that we consider the problem in two space dimensions with $u = 0$ and $\partial u/\partial t = v_0$ at $t = 0$, and use the two-dimensional Fourier transform

$$\hat{u}(k_1, k_2, t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, y, t)\mathrm{e}^{\mathrm{i}(k_1x+k_2y)}\,\mathrm{d}x\,\mathrm{d}y. \tag{4.63}$$

As in the three-dimensional case, we find that

$$\hat{u}(k, t) = \hat{v}_0(k, t)\frac{\sin a_0kt}{a_0k}, \tag{4.64}$$

where $k^2 = k_1^2 + k_2^2$. Following (4.36) and Exercise 4.8, and (4.60), we can write u in the form of a convolution integral if we can invert the function $\sin(a_0kt)/a_0k$. This immediately forces us into delicate convergence questions of the type mentioned in §4.4. One hair-raising possibility is to use the imaginary part of the following blind assertion: with $\rho^2 = x^2 + y^2$,

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}a_0kt}\mathrm{e}^{-\mathrm{i}(k_1x+k_2y)}}{a_0k}\,\mathrm{d}k_1\,\mathrm{d}k_2 &= \frac{1}{a_0}\int_0^{2\pi}\int_0^{\infty} \mathrm{e}^{\mathrm{i}k(a_0t-\rho\cos\theta)}\,\mathrm{d}k\,\mathrm{d}\theta \\ &= \frac{\mathrm{i}}{a_0}\int_0^{2\pi}\frac{\mathrm{d}\theta}{a_0t-\rho\cos\theta} \\ &= \begin{cases} 2\pi\mathrm{i}\left(a_0\sqrt{a_0^2t^2-\rho^2}\right)^{-1}, & \rho < a_0t, \\ 0, & \rho > a_0t. \end{cases}\end{aligned} \tag{4.65}$$

Squeamish readers may prefer to work backwards from the answer and check that the double Fourier transform of (4.65) is

$$\frac{2\pi\mathrm{i}}{a_0}\int_0^{2\pi}\int_0^{a_0t}\frac{\rho\mathrm{e}^{\mathrm{i}k\rho\cos\theta}}{\sqrt{a_0^2t^2-\rho^2}}\,\mathrm{d}\rho\,\mathrm{d}\theta, \tag{4.66}$$

which is an integral that can, with the use of tables [19], be reduced to $\mathrm{e}^{\mathrm{i}a_0kt}/a_0k$.

The discussion above, like that at the end of the preceding section, raises the general question of identifying how much of the region of influence of any compactly supported Cauchy data is actually excited when we are simply considering the wave equation (4.55). It is tempting to conjecture that in one and three dimensions the solution always has a sharp 'beginning' on the outgoing characteristic surfaces through the boundary of the support, and a sharp 'conclusion' according to Fig. 4.6; we get no such sharp conclusions in two dimensions. While, strictly speaking, this statement is true, it is important to remember that it only applies to the constant-coefficient wave equation (4.55). Any inhomogeneity such as a spatial variation in a_0 probably destroys the property that the conclusion of the disturbance is sharp.

We conclude this section by mentioning how Fourier transforms lead to yet another interesting representation of solutions of the wave equation. In all the cases we have considered with Cauchy data $u = 0$ and $\partial u/\partial t = v_0$ at $t = 0$,[46] we have found that the Fourier transform of u is $\hat{u} = \hat{v}_0(\sin a_0kt)/(a_0k)$. As described after (4.36) and in Exercise 4.8, this gives the explicit one-dimensional solution as

$$\begin{aligned} u(x,t) &= \int_{-\infty}^{\infty} \mathrm{e}^{-ikx}\hat{v}_0(k)\frac{\sin a_0kt}{a_0k}\,\mathrm{d}k \\ &= \frac{1}{2a_0}\left(V_0(x+a_0t) - V_0(x-a_0t)\right), \end{aligned} \tag{4.67}$$

where $V_0'(x) = v_0(x)$.

However, in two dimensions, instead of (4.65), we can write

$$\begin{aligned} u(x,y,t) = \int_0^{2\pi}\int_0^{\infty} \Big(&\mathrm{e}^{-ik(x\cos\phi+y\sin\phi)-ia_0kt} \\ &-\mathrm{e}^{-ik(x\cos\phi+y\sin\phi)+ia_0kt}\Big)\hat{F}(k,\phi)\,\mathrm{d}k\,\mathrm{d}\phi, \end{aligned}$$

where $k_1 = k\cos\phi$, $k_2 = k\sin\phi$ and $\hat{F} = \hat{v}_0/(4\pi^2a_0k)$. Thus, when we integrate with respect to k, we have a Fourier inversion with arguments $x\cos\phi + y\sin\phi \pm a_0t$. Hence,

$$u(x,y,t) = \int_0^{2\pi}\left(F(x\cos\phi + y\sin\phi + a_0t,\phi) - F(x\cos\phi + y\sin\phi - a_0t,\phi)\right)\mathrm{d}\phi. \tag{4.68}$$

While this 'plane-wave' superposition is less useful than (4.61) for solving initial value problems, it is interesting that it involves just *one* integral of an arbitrary function of two variables.

Finally, in three dimensions, the obvious generalisation of (4.68) is the plane-wave superposition

$$\begin{aligned} u(x,y,z,t) = \int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi} &\left(F(x\sin\phi\cos\theta + y\sin\phi\sin\theta + z\cos\phi + a_0t,\theta,\phi)\right. \\ &\left. - F(x\sin\phi\cos\theta + y\sin\phi\sin\theta + z\cos\phi - a_0t,\phi)\right)\mathrm{d}\phi\,\mathrm{d}\theta. \end{aligned} \tag{4.69}$$

4.6.2 Hyperbolicity and time-likeness

In §4.3 we have already made some remarks about the way in which the idea of time-like directions can be introduced mathematically into the theory of initial/boundary value problems in one space dimension in order to suggest well-posed data. In more space dimensions, such considerations have greater physical

[46] Remember we can consider more general initial value problems using the ideas of p. 121.

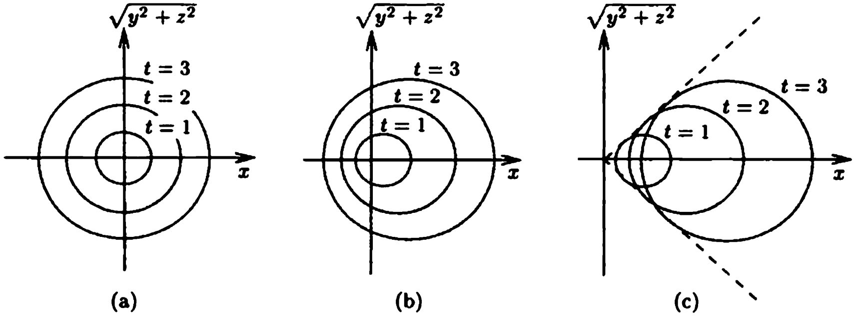


Fig. 4.7 Response to a localised source for (4.70) when (a) $U = 0$, (b) $0 < U < a_0$ and (c) $a_0 < U$.

significance, as can be seen by considering acoustics in a medium moving with speed U. When we linearise the equations of gas dynamics (generalised to three space dimensions) about the state $\mathbf{u} = U\mathbf{i}$, $p = p_0$ and $\rho = \rho_0$, as we did in §2.6, we find that the disturbance pressure, density and velocity potential all satisfy

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)^2 u = a_0^2\nabla^2 u. \tag{4.70}$$

Clearly, this equation can be reduced to (4.55) by the coordinate transformation $x - Ut = X$, but we propose to study it in a fixed frame of reference, with Cauchy data (4.57) prescribed at $t = 0$, with $v_0 = 0$ and u_0 localised near the origin. In three space dimensions when $U = 0$, this means that the sequence of spheres representing the evolution of the support of the response u is as given by

$$u = \frac{1}{r}F(r - a_0t),$$

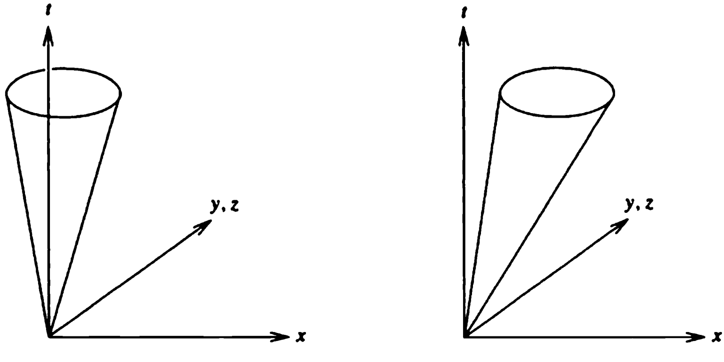
with F a delta function of argument $r - a_0t$ (see Fig. 4.7(a)). However, these spheres have their centres shifted in the x direction, as in Figs 4.7(b) and (c) for $0 < U < a_0$ and $U > a_0$, respectively. These pictures correspond to the effect of a 'burst' of sound from the origin, which we know is localised on the surface of successive spheres

$$(x - Ut)^2 + y^2 + z^2 = a_0^2t^2, \tag{4.71}$$

at successive intervals of time.[47]

There is clearly an important distinction between Figs 4.7(b) and (c), and it is one that we have anticipated in §4.3. In the first case, the time axis lies in the interior of (4.71), thought of as a manifold in four-dimensional space, which is what we called the ray cone in Chapter 2. Moreover, in this case, all points in

[47] If, instead of postulating an initial point source in Fig. 4.7(a), we were to postulate an initial 'wave' which was non-spherical, the subsequent evolution of that wave, which, as we will see in Chapter 8, can be traced by moving along the normal at each point with speed a_0, could easily lead to the front developing singularities.



(a) (b)

Fig. 4.8 Time-like and space-like ray cones for (4.70).

space eventually receive the acoustic signal. However, if $U > a_0$, the time axis lies outside the cone (4.71), as shown schematically in Fig. 4.8, and time (with (x, y, z) fixed) is no longer 'time-like'. In this latter case only those directions (x, y, z) inside the so-called *Mach cone*, namely

$$y^2 + z^2 < \frac{a_0^2 x^2}{U^2 - a_0^2} \quad \text{for } x > 0,$$

receive the signal.

This example illustrates the idea that the definition of hyperbolicity needs to say more than simply that 'the characteristics are real' if it is to have a good physical interpretation. There must also be some distinction made between different directions in the space of the independent variables, and in particular time-like directions need to be identified as those interior to the inner sheet of the ray cone. Hence we now formally define a linear scalar second-order partial differential equation with m independent variables as *hyperbolic* at a point if there is a real characteristic (ray) cone of dimension $m - 1$ through that point. This definition clearly accords with the ideas of §2.6.

We note that the example (4.70) can be used to illustrate Huygens' principle in a time-independent situation. Suppose we consider steady flow so that (4.70) becomes

$$U^2 \frac{\partial^2 u}{\partial x^2} = a_0^2 \nabla^2 u, \tag{4.72}$$

which is hyperbolic if $U > a_0$. From the representation (4.62) with an appropriate change of notation, we see that a point disturbance at $x = y = z = 0$ in a three-dimensional flow (say, for example, the flow past a small bullet in a wind tunnel) produces a disturbance both *on* and *inside* the Mach cone $y^2 + z^2 = a_0^2 x^2/(U^2 - a_0^2)$. The bullet leaves a broad wake, but this would not be the case for a two-dimensional flow (say, past a straight wire placed perpendicular to the stream along the z axis), whose presence would only be felt *on* the cone $y^2 = a_0^2 x^2/(U^2 - a_0^2)$.

Incidentally, the model (4.72) also enables us to give a physical interpretation of the bicharacteristics that were introduced in §2.6. The Mach cone is simply the characteristic surface across which any of the flow variables can have a jump in its second derivatives. Now suppose the bullet is replaced by the slender pointed nose of an aircraft model and that a thin aerial protrudes from the nose through the Mach cone. Since the aerial only intersects the cone at one point, it propagates a discontinuity just along the bicharacteristic through that point which, in this case, is just the generator of the Mach cone. Any 'fatter' object that protrudes through the Mach cone would generate a new component of the characteristic surface.

As a final remark to lead into the next section, we note that, had we considered an equation more complicated than $\partial^2 u/\partial t^2 = a_0^2 \nabla^2 u$, such as one with variable coefficients, then diagrams such as Fig. 4.7 would for small time become more complicated. In particular, the spherical 'waves' in Fig. 4.7 would locally become ellipsoids whose shape depended on the coefficients in the equation, and the ray cones would correspondingly have elliptical cross-sections. The geometric complexity would become much worse if the cones were to lose their local convexity; this cannot happen for scalar second-order equations, but it may for vector equations, as we shall soon see.

*4.7 Higher-order systems

One of the basic motivations for studying partial differential equations is that so many everyday phenomena are modelled by systems of such equations. Indeed, in the preceding sections we have explained how even a scalar hyperbolic equation can not only underpin much of the science of acoustics but also describe various other kinds of small-amplitude wave motion. We now consider two other areas of applied science which lead to vector second-order hyperbolic equations.

4.7.1 Linear elasticity

The first concerns the propagation of acoustic waves in solids, whose modelling involves some elementary elasticity theory. As is the case for all continuum mechanics, we introduce a vector field $\mathbf{u}(\mathbf{x}, t)$ to describe the configuration of the medium. In this case it represents the displacement of a particle from an unstressed reference configuration, i.e. the particle which is at $\mathbf{x}$ in the reference configuration moves by $\mathbf{u}(\mathbf{x}, t)$. We also need to model the forces in the medium, and this should be done by introducing a stress tensor (as, for example, in [40]). Here we side-step this process by merely noting that elastic solids resist both shear forces (Fig. 4.9(a)) and tensile/compressive forces (Fig. 4.9(b)).

An example of the former would be the shearing of a sheet of rubber with a displacement u in the x direction that is only a function of y and t. In the simplest model, we could propose that the shear force per unit area in the sheet is $\mu\, \partial u/\partial y$, where μ is the constant 'shear modulus', and hence an argument like that used in §2.1 gives that the equation of motion is

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial y}\left(\mu \frac{\partial u}{\partial y}\right),$$

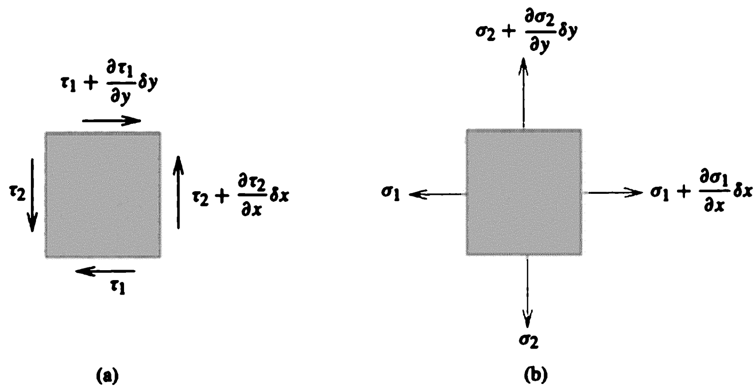


Fig. 4.9 (a) Shear stresses and (b) tensile stresses in an elastic medium.

where ρ is the density of the sheet. Equally, if we were to stretch a rubber band along its length so that its displacement is $u(x,t)$ in the x direction, the tensile force would be $\lambda' \, \partial u/\partial x$ and the equation of motion would be

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(\lambda' \frac{\partial u}{\partial x}\right),$$

where λ' is another positive constant. The effect of combining these two mechanisms in order to synthesise the displacement $\mathbf{u}(\mathbf{x},t)$ of a general isotropic linearly elastic solid is not quite a simple summation, because of the coupling between compression/tension and shear. Hence we simply quote the end result,

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = (\lambda + 2\mu)\nabla\nabla \cdot \mathbf{u} - \mu \nabla \wedge (\nabla \wedge \mathbf{u}), \tag{4.73}$$

where μ, the *shear modulus* of the material, is as above and λ is related to *Young's modulus* λ' by $\lambda' = \mu(3\lambda + 2\mu)/(\lambda + \mu)$.

We are thus confronted with a system of three linear second-order equations, but there is no reason why we cannot combine the approach used in §2.6 with that used earlier in this chapter: we simply define $\phi(x,y,z,t) = 0$ to be a characteristic surface if it is one on which the prescription of Cauchy data $\mathbf{u}$ and $\partial \mathbf{u}/\partial n$ does not yield a unique local solution. Writing

$$\left(\frac{\partial \phi}{\partial t}, \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial z}\right) = (\xi_0, \xi_1, \xi_2, \xi_3),$$

as in §2.6, we find, after some manipulation, that

$$\left(\mu(\xi_1^2 + \xi_2^2 + \xi_3^2) - \rho\xi_0^2\right)\left((\lambda + 2\mu)(\xi_1^2 + \xi_2^2 + \xi_3^2) - \rho\xi_0^2\right) = 0. \tag{4.74}$$

Hence, in the language of §2.6, the normal cone comprises two components rather than the single 'sheet' of fluid acoustics in §4.6.2 above. The corresponding ray cone also comprises two sheets. This is geometrically obvious from the fact that the normal cone has concentric circles as cross-sections $\xi_0 =$ constant, but we can also compute the ray cone analytically; from the prescription in (2.54)–(2.56), it is given parametrically by

$$\frac{x}{\mu} - \Lambda\xi_1 = \frac{y}{\mu} - \Lambda\xi_2 = \frac{z}{\mu} - \Lambda\xi_3 = \frac{t}{\rho} + \Lambda\xi_0 = 0,$$

and

$$\frac{x}{\lambda+2\mu} - \Lambda\xi_1 = \frac{y}{\lambda+2\mu} - \Lambda\xi_2 = \frac{z}{\lambda+2\mu} - \Lambda\xi_3 = \frac{t}{\rho} + \Lambda\xi_0 = 0,$$

i.e.

$$(x^2 + y^2 + z^2 - c_s^2 t^2)(x^2 + y^2 + z^2 - c_p^2 t^2) = 0, \tag{4.75}$$

where $c_s^2 = \mu/\rho$ and $c_p^2 = (\lambda+2\mu)/\rho$ are the so-called 'S' and 'P' wave speeds. The physical interpretation is that acoustic waves in an elastic solid can propagate at two distinct velocities, as is apparent from seismograph traces.[48]

This example shows us that, if we wish to generalise the arguments about hyperbolicity given at the end of §4.6 to higher-order systems, we must be prepared to consider 'multi-sheeted' normal and ray cones rather than the single-sheeted cones of Fig. 4.8. Fortunately, both the cones (4.74) and (4.75) consist of two 'nested' convex cones, and energy estimates like those mentioned in Exercise 4.13 can be used to prove well-posedness of the solution whenever Cauchy data is prescribed on a space-like surface, i.e. on one that lies 'outside' the outer sheet of the ray cone; this implies that the normal to a conical space-like surface always lies in a time-like direction inside the outer sheet of the ray cone. As in §4.6, this geometry provides a strong motivation for the definition of hyperbolicity for higher-order systems to be associated not only with the system of partial differential equations but also with the relevant time- and space-like directions in the space of the independent variables. However, the fact that (4.74) has two components now opens up the possibility that their union might not be convex, and we will return to this situation at the end of this section.

Finding solutions of (4.73) in practical situations entails many serious technical details, especially when it comes to applying boundary conditions. One common approach is to consider solutions in the form

$$\mathbf{u} = \Re\left(\mathbf{u}'(\mathbf{x})e^{-i\omega t}\right) \quad \text{for } \omega \in \mathbb{R},$$

where $\mathbf{u}'$ satisfies an equation in the so-called *frequency domain* with only three independent variables. However, this latter equation is no longer hyperbolic and will be dealt with in the next chapter. Whether we deal with systems in the real time domain or the frequency domain by transform methods, we inevitably encounter difficulties caused by having to consider the Fourier inversion of rational functions involving high-degree polynomials (sextic in this case). Hence we will not consider any explicit solutions in this chapter.

[48] In fact, earthquake recording is complicated by the presence of the stress-free boundary of the earth, which is responsible for generating a third wave speed, as shown in Exercise 4.16.

4.7.2 Maxwell's equations of electromagnetism

We conclude by mentioning perhaps the most famous of all higher-order systems of linear partial differential equations, Maxwell's equations of electromagnetism. Despite their importance, these equations are far more difficult to explain than any we have yet encountered. Hence, we will only give a brief account of their derivation. The starting point is *Faraday's law of induction*, which concerns the basic mechanism for dynamos and aerials. It says that any time-varying magnetic flux through a closed loop L produces an electric field $\mathbf{E}$ around L, and is easy to verify in a laboratory where, if L is a conductor, the electric field drives a current round it. Denoting the magnetic field by $\mathbf{H}$, the rate of change of magnetic flux through L, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\iint_D \mathbf{H}\cdot \mathrm{d}\mathbf{S} = \iint_D \frac{\partial \mathbf{H}}{\partial t}\cdot \mathrm{d}\mathbf{S},$$

where D spans L. If L happens to be a conductor, the electric field generated drives a current round L which, by Ohm's law, is itself proportional to

$$\int_L \mathbf{E}\cdot \mathrm{d}\mathbf{x} = \iint_D \nabla\wedge\mathbf{E}\cdot \mathrm{d}\mathbf{S}.$$

Thus, since D is arbitrary, $\partial\mathbf{H}/\partial t \propto \nabla\wedge\mathbf{E}$. The constant of proportionality turns out to be negative and the equation is usually written as

$$\nabla\wedge\mathbf{E} = -\mu\frac{\partial \mathbf{H}}{\partial t}. \tag{4.76}$$

However, the electric field is generated whether or not L is conducting and Maxwell proposed that this, the first of his equations, should hold everywhere and not just for conductors in which $\mathbf{E}$ is associated with a current.

The second, and even more intuitive, of Maxwell's equations comes from another experimental observation. This is the Biot–Savart law, which asserts that a current j flowing steadily along a straight wire creates a magnetic field $\mathbf{H}$ azimuthally around the wire and that its strength decays inversely with the square of the distance from the wire. Hence the magnetic field $\mathbf{H}(\mathbf{x})$ caused by any current loop L is proportional to

$$-j\int_L \frac{(\mathbf{x}-\mathbf{x}')\wedge \mathrm{d}\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|^3}; \tag{4.77}$$

in the usual units the constant of proportionality is $1/4\pi$. It is then plausible to propose that, for an arbitrary steady spatially-distributed current $\mathbf{j}$,

$$\nabla\wedge\mathbf{H} = \mathbf{j}, \tag{4.78}$$

but we will have to wait until Chapter 5 before we can see why (4.77) can be viewed as the 'solution' of (4.78).

The Biot–Savart law does not allow for the effects of time-varying currents or charge distributions but, if we assume conservation of charge, we must ensure that

$\mathbf{j}$ is related to the charge density ρ by an equation analogous to the conservation laws mentioned in §2.1. Thus,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \tag{4.79}$$

However, in the same way that the divergence of a gravitational force field is proportional to the local mass density, we can propose that

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon}, \tag{4.80}$$

where ϵ is a constant called the *permittivity*. Maxwell's great contribution was to assert that the effect of time dependence is to change (4.78) to

$$\nabla \wedge \mathbf{H} = \mathbf{j} + \epsilon \frac{\partial \mathbf{E}}{\partial t}, \tag{4.81}$$

thereby ensuring consistency with (4.79).[49]

We are not quite finished, as it can be seen from the partial differential equation point of view that an additional scalar equation for $\mathbf{H}$ is necessary for us to have any hope of uniqueness. To see this, let us restrict ourselves to free space where there is no charge and no current, so that

$$\nabla \wedge \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \quad \text{and} \quad \nabla \wedge \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}. \tag{4.82}$$

Although these are six equations for six unknowns, the degeneracy of the curl operator means that, using these equations alone, $\mathbf{E}$ and $\mathbf{H}$ can each only be determined to within the gradient of a scalar, time-independent function. However, (4.80) gives[50]

$$\nabla \cdot \mathbf{E} = 0, \tag{4.83}$$

which removes the non-uniqueness in $\mathbf{E}$, and a similar scalar equation is also necessary for $\mathbf{H}$. Fortunately, the consequence of the commonly-held belief that there are no magnetic monopoles is that

$$\nabla \cdot \mathbf{H} = 0, \tag{4.84}$$

which, unlike (4.83), holds even in the presence of charges or currents.

[49]We must warn the reader that (4.76) and (4.81) are usually derived much more fundamentally in physics texts in terms of the so-called magnetic induction $\mathbf{B}$ and displacement current $\mathbf{D}$ rather than $\mathbf{H}$ and $\mathbf{E}$, respectively. In many practical situations the distinction is fortunately irrelevant from the mathematical standpoint.

[50]We remark that $\nabla \cdot \mathbf{E} = 0$ is also true to a good approximation in many cases of steady current flow, for example in metals. In such situations we can often use the bulk version of Ohm's law, $\mathbf{j} = \sigma \mathbf{E}$, where σ is a material property called the electrical conductivity, to show from (4.79) and (4.80) that

$$\frac{\partial \rho}{\partial t} + \frac{1}{\epsilon}(\sigma \rho + \nabla \sigma \cdot \mathbf{E}) = 0;$$

when σ is constant, we must have $\rho = 0$ (indeed, any initial charge density decays on the time-scale ϵ/σ). It is only when σ has large spatial variations that charge can accumulate; an example of this is the surface charge layer that can occur when σ jumps between a conductor and an insulator.

Combining these equations and using the identity $\nabla \wedge \nabla \wedge = \nabla(\nabla \cdot) - \nabla^2$ gives that, in a vacuum,

$$\left(\frac{\partial^2}{\partial t^2} - c^2\nabla^2\right)\mathbf{H} = \left(\frac{\partial^2}{\partial t^2} - c^2\nabla^2\right)\mathbf{E} = \mathbf{0},$$

where $c^2 = (\epsilon\mu)^{-1}$, and hence the magnetic and electric fields both satisfy vector wave equations with speed of propagation c, the velocity of light. From previous arguments, both equations are hyperbolic with time-like directions in $x^2+y^2+z^2 < c^2t^2$ (the normal cone is $c^2(\xi_1^2 + \xi_2^2 + \xi_3^2) - \xi_0^2 = 0$) but, combined as a sixth-order system for $\mathbf{H}$ and $\mathbf{E}$, we find a mathematical degeneracy because the normal cone does not have distinct components.

The examples above from acoustics and electromagnetism give us an excellent physical framework in which to interpret characteristic surfaces in multi-variable partial differential equations. These 'manifolds' $\phi(x, y, z, t) = 0$ are, as we know, surfaces across which discontinuities in the second derivatives of the dependent variable can propagate. Hence, it is natural to think of them as 'wave-fronts', as we did in Fig. 4.8. Many people also have a good physical intuition about light rays and sound rays but it can sometimes be quite difficult to identify such rays with the generators of the ray cone in any time-dependent propagation problem. One difficulty, at which we have already hinted, is that the normal cone may not be nested or convex, as it was for linear elastic waves or electromagnetic waves in a vacuum. A commonly occurring configuration is that of two cones of elliptical cross-section, as in Fig. 4.10, and, indeed, this geometry can be shown to arise when a sufficiently anisotropic electromagnetic medium is modelled by replacing $\mathbf{E}$ in (4.82) by a suitable constant matrix multiplied by $\mathbf{E}$. Now suppose we try to calculate the ray cones in such a situation. The ray cones corresponding to the

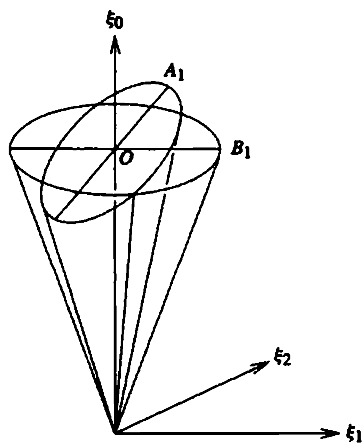


Fig. 4.10 Normal cones for an anisotropic medium; $\xi_0 = t$, $\xi_1 = x$ and $\xi_2 = y$.

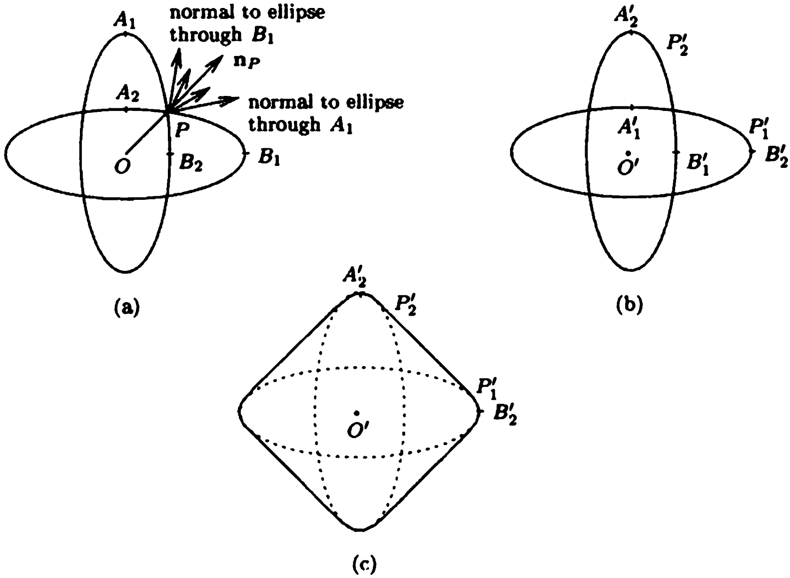


Fig. 4.11 (a) Cross-section ξ_0 = constant of normal cone; (b) formal t = constant cross-section of ray cone; (c) convex hull of t = constant cross-section of ray cone.

two elliptical components of the normal cones also have elliptical cross-sections, as shown in Fig. 4.11, where the normal OA_1 in Fig. 4.11(a) gives rise to the tangent at A_1' in Fig. 4.11(b), etc. However, a difficulty arises when we carry out the calculation in the direction $\mathbf{n}_p$, because this gives rise to *two* tangents, at P_1' and P_2'; indeed, the normals belonging to the normal cone at P comprise the fan of vectors shown in Fig. 4.11(a). The details of the calculation are tiresome (see Exercise 4.19), but they eventually reveal that each vector in this fan gives rise to the double tangent $P_1'P_2'$ in Fig. 4.11(c). Hence the non-convexity of the normal cone has created a new component of the ray cone and this construction of the 'convex hull' of the ray cone can be shown to be valid quite generally. Its physical interpretation is that no plane light wave could be propagated with a normal $\mathbf{n}_p$ to its wave-fronts; instead the light would spread out in all the directions between P_1' and P_2'.

This concludes our discussion of linear hyperbolic equations for the moment. There is a great deal more that can be said about such equations, even at quite a general level, and, in particular, we would cite the important practical issue of what constitute well-posed boundary conditions at interfaces between two adjacent regions in each of which a hyperbolic equation is to be satisfied. For example, it is vital that the correct components of the relevant solution vectors are continuous at an interface between two different elastic solids or two different electromagnetic media. A good rule of thumb is to ensure that when the hyperbolic equation is written just in terms of the first derivatives of physically relevant quantities, the 'integration by parts' or 'weak solution' identities such as (2.43) are satisfied. An illustration of this can be found in Exercise 4.17.

4.8 Nonlinearity

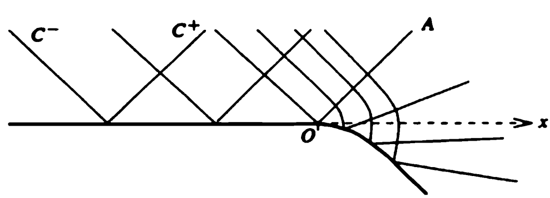
We have already seen in Chapter 1 that the presence of nonlinearity usually results in fundamental changes in the properties of solutions of previously linear partial differential equations. While, as a general rule, every nonlinear equation needs to be treated on its own merits, there are just a few general statements that can be made and general techniques that can be tried. Of course, when these techniques apply, they are worth their weight in gold.

There is one approach that is peculiar to hyperbolic equations, where we have seen that a common effect of nonlinearity is that of shock formation as a result of characteristics intersecting. But, before such catastrophes occur, there may be the following possibilities.

4.8.1 Simple waves

We remarked in Chapter 2 that characteristics are usually of little value in obtaining explicit solutions of hyperbolic systems. Even for a scalar second-order quasilinear equation with two independent variables, it is very unlikely that we can find Riemann invariants, i.e. functions of x, y, u, $\partial u/\partial x$ and $\partial u/\partial y$ that are constant on a family of characteristics. Moreover, when such a happy circumstance does occur, the initial and boundary conditions almost certainly do not prescribe the values of these invariants. However, one situation exists where we can always make progress, and it occurs surprisingly often in practice: it is when we consider the effect of a sudden change in the boundary data on a solution that is trivial before the change occurs. A prototypical example is that of two-dimensional gas flow with speed U_∞ and Mach number $M_\infty > 1$ past an initially straight wall (Figs 4.12 and 4.13). The relevant nonlinear hyperbolic equation is

(a)



(b)

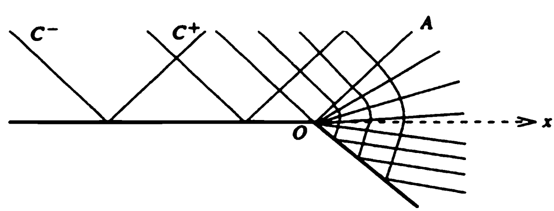


Fig. 4.12 (a) Smooth and (b) abrupt expansion of a gas flow.

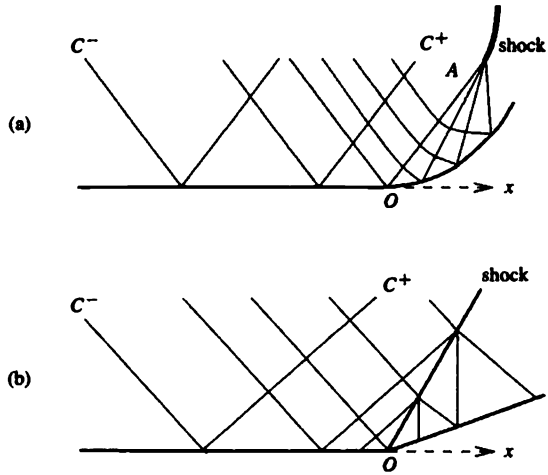


Fig. 4.13 (a) Smooth and (b) abrupt compression of a gas flow.

$$F\frac{\partial^2 u}{\partial x^2} + G\frac{\partial^2 u}{\partial x \partial y} + H\frac{\partial^2 u}{\partial y^2} = 0, \tag{4.85}$$

where F, G and H depend on $\partial u/\partial x$ and $\partial u/\partial y$ only, but their precise form does not concern us here. All that matters is that there should exist Riemann invariants, i.e. functions of $\partial u/\partial x$ and $\partial u/\partial y$ that are constant on each family of characteristics $C^{\pm}$. The undisturbed unidirectional flow in $x < 0$ gives known, and in fact constant, values for $\partial u/\partial x$ and $\partial u/\partial y$ there, so that the characteristics are straight. Additionally, these values of $\partial u/\partial x$ and $\partial u/\partial y$ give Cauchy data on the non-characteristic $x = 0$ and this Cauchy problem must be solved in conjunction with the change of boundary conditions that occurs at $x = 0$, $y = 0$. To ensure uniqueness of the solution, we must also insist that there is no 'upstream influence', so that the region of influence of the origin is to the right of the characteristics OA in Figs 4.12 and 4.13.

Now the key observation is that, since the Riemann invariants are explicitly known as functions of $\partial u/\partial x$ and $\partial u/\partial y$, and since they are constant everywhere in the undisturbed flow, then, assuming no jumps occur in these variables, one Riemann invariant is constant in the disturbed flow, at least until any characteristics of the same family start to intersect. Let us look first at the configuration of Fig. 4.12, where the Riemann invariant on the characteristics C^- is constant. Hence, at least to begin with, the disturbed solution is simply given by solving a *first-order* scalar equation; we have already seen an example of this in Exercise 2.13. Moreover, the constancy of the Riemann invariant on C^- means that the characteristics C^+, on which a second function of $\partial u/\partial x$ and $\partial u/\partial y$ is constant, are straight, and in this case they can be determined explicitly in terms of the boundary conditions.

This is the typical situation in *simple wave flow*. The adjacency of the region of interest to a region in which the solution has known constant values for its Riemann invariants means that the order of the problem can be reduced, which often means that it can be solved explicitly.

In either the smooth or sharp geometries in Fig. 4.12, the C^+ characteristics form what is called an *expansion fan*. However, when the geometry is as in Fig. 4.13, these C^+ characteristics converge and eventually a shock forms; in the extreme case of an abrupt corner, as in Fig. 2.14, this shock is initiated at the corner.

The configuration of Fig. 4.12(b) is interesting because the characteristics in the fan are $y/x =$ constant and we find that ∇u only depends on y/x, not on x or y independently. Following the derivation of (4.15), this is our second encounter with what we will call a similarity solution in Chapter 6, because the solution at (x, y) is effectively the same as that at $(\lambda x, \lambda y)$ for any constant λ. We will see later that, if such an invariance exists, it can always be used to find certain solutions of a partial differential equation in m independent variables as solutions of a related equation with $m - 1$ independent variables.

Finally, it is interesting to note that combinations of straight shocks and expansion fans of the type described above, so-called *N-waves*, are sometimes found to describe the asymptotic behaviour of the solution of Cauchy problems for general hyperbolic equations. For example, it can be shown that as $t \to \infty$ the solution of

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0 \quad \text{for } -\infty < x < \infty,\ t > 0,$$

with $u(x, 0) = u_0(x)$ and f convex (i.e. $\mathrm{d}^2 f/\mathrm{d}x^2 > 0$), tends to an N-wave as $t \to \infty$, with the number of shocks being related to the number of maxima of u_0.

Unfortunately, gasdynamic flows involving interacting simple waves, or more general partial differential equations with three or more independent variables, rarely admit such simple solutions. However, gas dynamics suggests one other powerful technique which applies to some more general problems, hyperbolic or otherwise.

4.8.2 Hodograph methods

Suppose we have a second-order scalar hyperbolic equation in two variables (such as the transonic flow model (3.12)) whose coefficients depend only on the first derivatives of u. It is natural to consider a change of independent variables to $(\partial u/\partial x, \partial u/\partial y)$, but what would be the best choice of a dependent variable in such a case?

We can answer this question geometrically. If we think of the solution $u(x, y)$ as a surface S in (u, x, y) space and ask how to define w so that the surface $w = w(\partial u/\partial x, \partial u/\partial y)$ in $(w, \partial u/\partial x, \partial u/\partial y)$ space is as closely related to S as possible, we recall 'geometric duality'. In this concept, instead of only regarding a surface as a collection of points satisfying $u = u(x, y)$, we equally regard it as the envelope of its tangent planes

$$(x - x_0)\frac{\partial u}{\partial x}(x_0, y_0) + (y - y_0)\frac{\partial u}{\partial y}(x_0, y_0) = u - u_0,$$

as (x_0, y_0) vary with $u_0 = u(x_0, y_0)$. Hence, writing $p = \partial u/\partial x$ and $q = \partial u/\partial y$ in order to save ink, if we think of x and y as functions of p and q and define

$$w(p,q) = px(p,q) + qy(p,q) - u(x(p,q), y(p,q)), \tag{4.86}$$

then this tangent plane is

$$w = w_0 = p(x_0, y_0)x_0 + q(x_0, y_0)y_0 - u_0.$$

Thus, with w defined by (4.86), the surface $w = w(p,q)$ defined pointwise in (w, p, q) space is the same surface as $u = u(x, y)$ in (u, x, y) space, and w is the variable that we should work with.

Equation (4.86) is called the *Legendre transform* of u and it means that the formula for the tangent planes to $u = u(x,y)$ in (x, y, u) space involves no differentiations in w space, and vice versa. Partial differential equations whose solution surfaces possess this property are clearly promising candidates for explicit solutions if we formulate them correctly. To do this we need the identity

$$\frac{\partial}{\partial x} = \frac{\partial p}{\partial x}\frac{\partial}{\partial p} + \frac{\partial q}{\partial x}\frac{\partial}{\partial q},$$

so that

$$\frac{\partial w}{\partial p} = x + p\frac{\partial x}{\partial p} + q\frac{\partial y}{\partial p} - \frac{\partial u}{\partial p} = x, \qquad \frac{\partial w}{\partial q} = y.$$

Now let us carry this further. Since

$$1 = \frac{\partial p}{\partial x}\frac{\partial^2 w}{\partial p^2} + \frac{\partial q}{\partial x}\frac{\partial^2 w}{\partial p\,\partial q}, \qquad 0 = \frac{\partial p}{\partial x}\frac{\partial^2 w}{\partial p\,\partial q} + \frac{\partial q}{\partial x}\frac{\partial^2 w}{\partial q^2},$$

we have the easy relationships

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{\Delta}\frac{\partial^2 w}{\partial q^2}, \quad \frac{\partial^2 u}{\partial x\,\partial y} = -\frac{1}{\Delta}\frac{\partial^2 w}{\partial p\,\partial q}, \quad \frac{\partial^2 u}{\partial y^2} = \frac{1}{\Delta}\frac{\partial^2 w}{\partial p^2},$$

where the discriminant

$$\Delta = \frac{\partial^2 w}{\partial p^2}\frac{\partial^2 w}{\partial q^2} - \left(\frac{\partial^2 w}{\partial p\,\partial q}\right)^2 = \left(\frac{\partial^2 u}{\partial x^2}\frac{\partial^2 u}{\partial y^2} - \left(\frac{\partial^2 u}{\partial x\,\partial y}\right)^2\right)^{-1}. \tag{4.87}$$

Equations such as (4.85) are eminently suitable for this kind of change of variables. Indeed, this is the reason that the method is so-called because in mechanics the *hodograph* is the plane of velocities. The ensuing linear equation is

$$F(p,q)\frac{\partial^2 w}{\partial q^2} - G(p,q)\frac{\partial^2 w}{\partial p\,\partial q} + H(p,q)\frac{\partial^2 w}{\partial p^2} = 0,$$

and various further simplifications are sometimes possible.

The most important realisation provoked by this gasdynamic example is that Δ, as defined by (4.87), is infinite in the following cases.

1. In a uniform stream, the whole of the regions above and to the left of OA in Figs 4.12 and 4.13 map into a single point $(U_\infty, 0)$ in the hodograph plane.
2. In a simple wave, such a region maps into the curve in the hodograph plane corresponding to the constancy of the relevant Riemann invariant.

These observations prompt us to ask the question 'when can the hodograph transformation go wrong?' Trouble will only come if Δ is zero or infinite, preventing the map from (x, y) to (p, q) from being one-to-one, and we can make two remarks about this. First, the equation $\Delta = \infty$ has the geometric interpretation that the surface $u = u(x, y)$ $(w = w(p, q))$ is *developable*; this means that it is generated by the tangent lines to some curve Γ or, more technically, it has zero Gaussian curvature.[51] Hence, when $\Delta = \infty$, it is not surprising that there are many points on the developable where the normal vectors $(p, q, -1)$ are identical, because the tangent plane to $u = u(x, y)$ everywhere along a generator of the developable is just the 'osculating plane' at the point where the generator touches Γ. Second, the equation $\Delta = 0$ is itself an interesting non-quasilinear equation for $w(p, q)$, to which we will return in Chapter 8.

Although the hodograph method can be applied equally to elliptic and parabolic equations, the remarks above indicate that it must be used with caution. We must also draw the reader's attention to the sad fact that, in practice, the dependence of the physical boundary conditions on (x, y) often makes it impossible to find appropriate boundary conditions for w. The principal exception in fluid dynamics is when the flow direction is known to be constant on the boundary, i.e. when the boundary comprises only straight segments, and we will return to this situation in Chapter 7.

4.8.3 Liouville's equation

There is one other device that we cannot resist mentioning here because of its applicability to an elliptic equation as well as a hyperbolic one. It concerns *Liouville's equation*

$$\frac{\partial^2 u}{\partial x\, \partial y} = e^u,$$

for which the apparently retrograde step of differentiating with respect to x uncovers an exact derivative with respect to y, namely

$$\frac{\partial}{\partial y}\left(\frac{\partial^2 u}{\partial x^2} - \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2\right) = 0.$$

We are amply repaid because we see that $\partial u/\partial x$ satisfies the *ordinary* differential (Ricatti) equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2 = f(x)$$

[51] A developable is a special case of a ruled surface, i.e. one formed by glueing together a one-parameter family of straight lines.

for some function f. Hence, by writing[52]

$$\frac{\partial u}{\partial x} = -\frac{2}{v}\frac{\partial v}{\partial x},$$

so that $u(x,y) = 2\log(F(y)/v(x,y))$ for an arbitrary function $F(y)$, we have reduced Liouville's equation to the linear problem

$$\frac{\partial^2 v}{\partial x^2} + \frac{1}{2}f(x)v = 0.$$

Although this gives the general solution, containing as it does the two arbitrary functions $f(x)$ and $F(y)$, the dependence of u on f and F is far from transparent. However, we can make more progress by noting that we could equally well have differentiated first with respect to y, giving that $u(x,y) = 2\log(G(x)/w(x,y))$, where

$$\frac{\partial^2 w}{\partial y^2} + \frac{1}{2}g(y)w = 0,$$

for two more arbitrary functions $g(y)$ and $G(x)$. The final coup is to exploit the structure of solutions of second-order linear equations, writing

$$v(x,y) = a_1(y)v_1(x) + a_2(y)v_2(x), \qquad w(x,y) = b_1(x)w_1(y) + b_2(x)w_2(y),$$

so that, since $F(y)/v(x,y) = G(x)/w(x,y)$,

$$\frac{G(x)v_i(x)}{b_i(x)} = \gamma = \frac{F(y)w_i(y)}{a_i(y)} \quad \text{for } i = 1,2,$$

where γ is constant. But the constancy of the Wronskians of the equations for v and w shows that we can take

$$v_2(x) = v_1(x)\int^x \frac{ds}{v_1^2(s)}, \qquad w_2(y) = w_1(y)\int^y \frac{ds}{w_1^2(s)},$$

from which there are functions $X(x)$ and $Y(y)$ such that

$$v_1 = \frac{1}{\sqrt{X'}}, \quad v_2 = \frac{X}{\sqrt{X'}}, \qquad w_1 = \frac{1}{\sqrt{Y'}}, \quad w_2 = \frac{Y}{\sqrt{Y'}},$$

and so

$$\frac{F}{v} = \frac{\gamma}{w_1v_1 + w_2v_2} = \gamma\frac{\sqrt{X'(x)Y'(y)}}{1 + X(x)Y(y)}.$$

The final step is to take $\gamma = \sqrt{2}$, so that the general solution of Liouville's equation is given in terms of two arbitrary real functions $X(x)$ and $Y(y)$ as

$$u(x,y) = \log\left(\frac{2X'(x)Y'(y)}{(1 + X(x)Y(y))^2}\right).$$

Other choices of γ give the general solution of $\partial^2 u/\partial x\,\partial y = ce^u$ for $c \neq 1$.

[52]The rationale for this transformation will be explained in Chapter 6.

*4.8.4 Another method

Apart from the many other ingenious tricks that exist in the literature, but are too specialised to list here, there is one other 'heavy gun' that can sometimes be wheeled out when dealing with nonlinear hyperbolic equations, and that is *soliton theory* [15]. This method is complicated technically and, since some of its most famous successes concern non-hyperbolic equations, we will defer discussion until Chapter 9. Meanwhile, readers might like to consider what they would do if faced with the Sine–Gordon equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = \sin u,$$

which arises in superconductivity modelling.

Exercises

4.1. Verify the result in footnote 39 by showing that, if $\xi = ax + by$ and $\eta = cx + dy$, then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/\epsilon^2}\, d\xi\, d\eta = \pi\epsilon^2 \left|\det\begin{pmatrix} a & b \\ c & d\end{pmatrix}\right|.$$

4.2. Show that, if $\mathcal{L} = \partial^2/\partial t^2 - \partial^2/\partial x^2$ and homogeneous Cauchy data $u = \partial u/\partial t = 0$ is given on $t = 0$, then D in Fig. 4.1 is the triangle with vertices at (x,t), $(x-t,0)$ and $(x+t,0)$, and $R = 1/2$. Hence show that the solution of $\mathcal{L}u = e^{-x-t}$ in $t > 0$ is

$$u(x,t) = \int\int_D e^{-\xi-\tau}\, d\xi\, d\tau = \frac{1}{4}\left((2t+1)e^{-x-t} - e^{-x+t}\right).$$

4.3. The function u satisfies

$$\frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial x^2} = 1$$

almost everywhere in $x > 0$, $y > 0$, with $u(x,0) = \partial u/\partial y\,(x,0) = 0$ for $x > 0$, and $u(0,y) = 0$ for $y > 0$.
By extending u to be an odd function in x, so that $u(-x,y) = -u(x,y)$, and using a Riemann function, or otherwise, show that

$$u = \begin{cases} y^2/2, & x \geqslant y, \\ x(2y-x)/2, & y \geqslant x. \end{cases}$$

Verify that u and its first derivatives are continuous and evaluate the jumps in its second derivatives across the characteristic $y = x$.
Why did we say that the equation is satisfied only 'almost everywhere' and not everywhere?

4.4. Show that, if

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0,$$

then the Riemann invariants

$$\frac{\partial u}{\partial x} \pm \frac{\partial u}{\partial y}$$

are constant on the characteristics $y \pm x = \text{constant}$.
Suppose u satisfies Cauchy data

$$u = \frac{1}{2}(1 - 2xy), \quad \frac{\partial u}{\partial r} = 1 - 2xy$$

on the semicircle $r^2 = x^2 + y^2 = 1$, $y < 0$. Show that a solution exists in the form $a(x-y)^2 + b(x+y)^2 + c$, where a, b and c are constant and that, for this solution, $\partial u/\partial x + \partial u/\partial y = 0$ everywhere and $\partial u/\partial x - \partial u/\partial y = \text{constant}$ on $x - y = \text{constant}$. However, the problem is ill-posed because the boundary is tangent to characteristics at $(\pm 1/\sqrt{2}, -1/\sqrt{2})$. Show non-existence when the data is perturbed so that

$$\frac{\partial u}{\partial x} - \frac{\partial u}{\partial y} \not\equiv 2(x - y)$$

on the arc between $(1/\sqrt{2}, -1/\sqrt{2})$ and $(1, 0)$.

4.5. Suppose that

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0$$

almost everywhere in $y > 0$, with

$$u = \frac{\partial u}{\partial y} = 0 \quad \text{on } y = 0,\ x > 0, \qquad u = 0, \quad \frac{\partial u}{\partial y} = 1 \quad \text{on } y = 0,\ x < 0.$$

Use a Riemann function to show that

$$u \equiv 0 \quad \text{in } x > y > 0, \qquad u \equiv y \quad \text{in } -x > y > 0.$$

By computing the Riemann invariants show that

$$\frac{\partial u}{\partial y} = -\frac{\partial u}{\partial x} = \frac{1}{2}$$

in $-y < x < y$ and check that u, but not its first derivatives, is continuous across $x = \pm y$. Confirm these results using d'Alembert's solution.

4.6. Suppose that
$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \delta(x - \tfrac{1}{2}t^2)$$
for $t > 0$, with $u(x, 0) = 0$. Show that putting $\xi = x - \frac{1}{2}t^2$ gives
$$(1 - t)\frac{\partial u}{\partial \xi} + \frac{\partial u}{\partial t} = \delta(\xi),$$
with $u(\xi, 0) = 0$. Deduce that, for $t < 1$, $u = 0$ for $x > t$ and $x < \frac{1}{2}t^2$. Show further that
$$u(x, t) \to \frac{1}{1 - t} \quad \text{as } x \downarrow \tfrac{1}{2}t^2,$$
and hence that
$$u(x, t) = \frac{1}{\sqrt{1 + 2(x - t)}}$$
for $\frac{1}{2}t^2 < x < t < 1$.
By considering the characteristics in the (ξ, t) plane, show that this solution holds for $t < 2$ and that, for $t > 2$,
$$u(x, t) = \begin{cases} 0, & x < t \text{ and } x > \frac{1}{2}t^2, \\ -1/\sqrt{1 + 2(x - t)}, & \frac{1}{2}t^2 < x < t. \end{cases}$$
Remark. Note that it would be easy to get the wrong answer to this question without the substitution $\xi = x - \frac{1}{2}t^2$. It is good technique to use the argument of a delta function as an independent variable where possible.

*4.7. A problem in the dynamics of the overhead power wire for an electric locomotive leads to the model
$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{for } x \neq X(t),\ t > 0,$$
where X is a prescribed smooth function with $0 < X'(t) < 1$ (t is time, $X(t)$ is the locomotive position and u is the displacement of the wire). Across $x = X$, there are prescribed discontinuities
$$\left[\frac{\partial u}{\partial x}\right]_{X-0}^{X+0} = -V(X(t), t), \quad [u]_{X-0}^{X+0} = 0.$$
Suppose $u = \partial u/\partial t = 0$ at $t = 0$. By modifying the argument leading to (4.8) to account for the discontinuities across $x = X$, show that
$$u(x, t) = \int \frac{1}{2}V(\xi, \tau)\left(1 - (X'(\tau))^2\right) d\tau,$$
where the integral is taken along the part of $\xi = X(\tau)$ that lies within $\tau - t < \xi - x < t - \tau$, $\tau > 0$. If $X = t^2/2$ and $V = 1$, show that, for $0 < x < t < \sqrt{2x}$,
$$u = \frac{1}{2}\left(\zeta - \frac{\zeta^3}{3}\right), \quad \text{where} \quad x - \frac{\zeta^2}{2} = t - \zeta,$$
and comment on the well-posedness of this situation.

4.8. Show that the Fourier transform in x of $\delta(x \pm a_0 t)$ is $e^{\mp i a_0 t}$. Now recall that the Fourier transform of the general solution of the one-dimensional wave equation (4.35) on p. 114 is

$$\hat{u}(k,t) = \hat{f}(k)e^{ika_0t} + \hat{g}(k)e^{-ika_0t}.$$

Use the convolution theorem to derive the general solution itself in the form

$$u(x,t) = f(x + a_0t) + g(x - a_0t).$$

4.9. (i) Suppose that $f(x) = 0$ in $x < 0$ and $F(x) = f(x)e^{-\beta x} \to 0$ as $x \to +\infty$ for some $\beta > 0$. Define

$$\hat{F}(k) = \int_{-\infty}^{\infty} F(x)e^{ikx}\,dx$$

and, assuming $\hat{F}$ can be suitably analytically continued, define $\hat{f}$ by

$$\hat{f}(k) = \hat{F}(k - i\beta).$$

Show that the inversion formula

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{F}(k)e^{-ikx}\,dk$$

becomes

$$f(x) = \frac{1}{2\pi}\int_{-\infty+i\beta}^{\infty+i\beta} \hat{f}(\kappa)e^{-i\kappa x}\,d\kappa.$$

Show also that this implies that the Laplace inversion formula is

$$f(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \tilde{f}(p)e^{px}\,dp,$$

where γ is such that all the singularities of $\tilde{f}$ lie in $\Re p < \gamma$.

(ii) Suppose that

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

for $t > 0$, with

$$u(x,0) = 0, \qquad \frac{\partial u}{\partial t}(x,0) = \begin{cases} 0, & x < 0, \\ e^{\alpha x}, & x > 0, \end{cases}$$

for some $\alpha > 0$. Show that

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty+i\beta}^{\infty+i\beta} \frac{(-\sin\kappa t)e^{-i\kappa x}\,d\kappa}{\kappa(\alpha + i\kappa)}$$

for $\beta > \alpha > 0$, and hence use the calculus of residues to show that

$$u(x,t) = \begin{cases} 0, & x < -t, \\ (e^{\alpha(x+t)} - 1)/2\alpha, & -t < x < t, \\ (e^{\alpha(x+t)} - e^{\alpha(x-t)})/2\alpha, & t < x. \end{cases}$$

4.10. Suppose that $f(x)$ satisfies Airy's equation

$$\frac{d^2 f}{dx^2} + xf = 0.$$

Verify that a solution is

$$f(x) = \int_{\tilde{\Gamma}} e^{-\zeta x} F(\zeta)\, d\zeta,$$

where $F(\zeta) = e^{-\zeta^3/3}$ and $\tilde{\Gamma}$ is such that the integral exists and the change in $e^{-\zeta^3/3 - x\zeta}$ from one end of $\tilde{\Gamma}$ to the other is zero. Find two choices of $\tilde{\Gamma}$ that cannot be deformed into each other.

4.11. (i) Take the Laplace transform of Bessel's equation of order zero,

$$r\frac{d^2 J_0}{dr^2} + \frac{dJ_0}{dr} + rJ_0 = 0,$$

and use the facts that $J_0(r)$ is analytic at $r = 0$ and $\int_0^\infty J_0(r)\, dr = 1$ to show that the Laplace transform of J_0 is

$$\overline{J}_0(p) = \frac{1}{\sqrt{1+p^2}}.$$

(ii) Show that, if Γ is such that differentiation under the integral sign is permissible, then

$$\int_\Gamma e^{r(t-1/t)/2} t^{-\nu-1}\, dt$$

satisfies Bessel's equation of order ν,

$$\frac{d^2 J_\nu}{dr^2} + \frac{1}{r}\frac{dJ_\nu}{dr} + \left(1 - \frac{\nu^2}{r^2}\right) J_\nu = 0.$$

Deduce that the Laplace transform of J_ν is proportional to

$$\frac{1}{\sqrt{1+p^2}} \frac{1}{\left(p + \sqrt{1+p^2}\right)^\nu}.$$

* 4.12. (i) Show that, if $\sqrt{1+k^2}$ is defined so that it asymptotes to k as $|k| \to \infty$ in all directions in the Argand diagram, a branch cut must be introduced between $k = \pm i$.

(ii) When we take $a_0 = c = 1$, the inversion of (4.51) is

$$u(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} v_0(\xi) e^{ik\xi}\, d\xi \right) \frac{\sin\left((t\sqrt{1+k^2}\right)}{\sqrt{1+k^2}} e^{-ikx}\, dk.$$

With the definition of $\sqrt{1+k^2}$ given above, reverse the order of integration and consider

$$\int_{-\infty}^{\infty} \frac{e^{ik(\xi - x) \pm it\sqrt{1+k^2}}\, dk}{\sqrt{1+k^2}}$$

along the contour of Fig. 4.14, inside which there are no singularities.

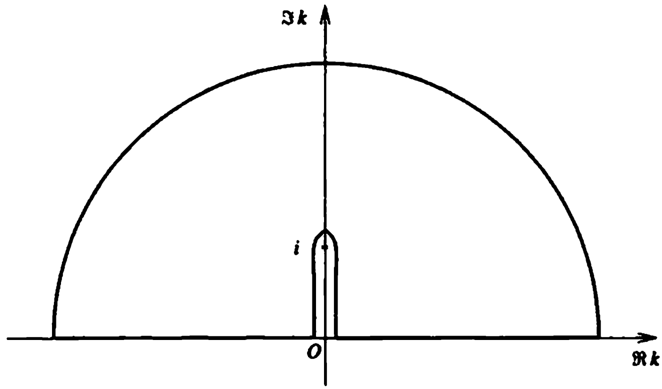


Fig. 4.14 Contour for Exercise 4.12.

Show that

$$\int_{-\infty}^{\infty} \frac{\sin\left(t\sqrt{1+k^2}\right) e^{ik(\xi-x)}}{\sqrt{1+k^2}}\,dk$$

$$= \begin{cases} 0, & \xi < x-t, \\ 2\displaystyle\int_0^1 \frac{\cosh(s\,(x-\xi))\cos\left(t\sqrt{1-s^2}\right)}{\sqrt{1-s^2}}\,ds, & x-t<\xi<x+t, \\ 0, & x+t<\xi, \end{cases}$$

with the non-zero contribution coming from the integrals along either side of the branch cut. Finally, quote the result that

$$\int_0^1 \cosh(as)\frac{\cos\left(b\sqrt{1-s^2}\right)}{\sqrt{1-s^2}}\,ds = \frac{\pi}{2}J_0\left(\sqrt{b^2-a^2}\right),$$

which follows from (4.47) on p. 117, to retrieve the Riemann function solution

$$u(x,t) = \frac{1}{2}\int_{x-t}^{x+t} J_0\left(\sqrt{t^2-(x-\xi)^2}\right) v_0(\xi)\,d\xi.$$

Check that, if we had defined $\sqrt{1+k^2}$ to tend to k if $\Re k > 0$ and $-k$ if $\Re k < 0$ as $|k| \to \infty$, we would obtain a domain of dependence of (x,t) that included points $(\xi, 0)$ such that $|\xi - x| > \tau$. Why would this be physically unrealistic?

4.13. Suppose that

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

in all space with $u = \partial u/\partial t = 0$ at $t = 0$, and that u decays sufficiently fast at infinity for

$$\iiint_{\mathbf{R}^3} \nabla u \cdot \nabla \left(\frac{\partial u}{\partial t} \right) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

to be integrated by parts. Show that the energy

$$\frac{1}{2} \iiint_{\mathbf{R}^3} \left(\left(\frac{\partial u}{\partial t} \right)^2 + |\nabla u|^2 \right) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

is identically zero and hence that the Cauchy problem (4.55) and (4.57) has a unique solution.

4.14. Show that, if

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \text{for } 0 < x, y < \pi,$$

with $u = 0$ on $x = 0$, π and on $y = 0$, π, and $u = 0$ and $\partial u / \partial t = 1$ at $t = 0$, then

$$u = \frac{16}{\pi^2} \sum_{m,n=0}^{\infty} \frac{\sin(2m+1)x \sin(2n+1)y \sin\left(t\sqrt{(2m+1)^2 + (2n+1)^2}\right)}{(2m+1)(2n+1)\sqrt{(2m+1)^2 + (2n+1)^2}}.$$

Show also that, if the boundary conditions are $u = 0$ in the triangle bounded by $y = 0$, $x = \pi$ and $x = y$, then there is a formal solution

$$u = \sum a_{nm} (\sin nx \sin my - \sin mx \sin ny) \sin\left(t\sqrt{n^2 + m^2}\right).$$

4.15. Show that, if $u(x, y, t)$ satisfies the same equation and boundary conditions as in the first part of Exercise 4.14, and that if

$$u(x, y, 0) = \sum_{m,n=1}^{\infty} a_{mn} \sin mx \sin ny,$$

$$\frac{\partial u}{\partial t}(x, y, 0) = \sum_{m,n=1}^{\infty} b_{mn} \sin mx \sin ny,$$

then its Laplace transform is

$$\tilde{u}(x, y, p) = \sum_{m,n=1}^{\infty} \frac{p\, a_{mn} + b_{mn}}{p^2 + m^2 + n^2} \sin mx \sin ny.$$

Hence retrieve the solution of Exercise 4.14.

*4.16. Suppose $\mathbf{u}$ satisfies the elasticity equations (4.73) and is of the form

$$\mathbf{u} = \Re\left((u_1(x, y), u_2(x, y), 0)^{\mathsf{T}} \mathrm{e}^{-\mathrm{i}\omega t} \right) \quad \text{for } y < 0.$$

Suppose also that

$$\mu \left(\frac{\partial u_1}{\partial y} + \frac{\partial u_2}{\partial x} \right) = 2\mu \frac{\partial u_2}{\partial y} + \lambda \left(\frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} \right) = 0$$

on $y = 0$, which corresponds to the traction-free boundary of a half-space (the earth, say). Noting the fact that, if $\Re\, \mathrm{e}^{\mathrm{i}(k_1 x + k_2 y - \omega t)}$ is a solution of (4.73),

(k_1, k_2, ω) lies in the normal cone $k_1^2 + k_2^2 = \omega^2/c^2$, use (4.75) to check that, for any constants A and B,

$$u_1 = kAe^{ikx+y\sqrt{k^2-\omega^2/c_p^2}}, \quad u_2 = -iA\sqrt{k^2 - \frac{\omega^2}{c_p^2}}\, e^{ikx+y\sqrt{k^2-\omega^2/c_p^2}}$$

and

$$u_1 = iB\sqrt{k^2 - \frac{\omega^2}{c_s^2}}\, e^{ikx+y\sqrt{k^2-\omega^2/c_s^2}}, \quad u_2 = kBe^{ikx+y\sqrt{k^2-\omega^2/c_s^2}}$$

are both candidate solutions of (4.73). Show further that the boundary conditions are satisfied if $k = \omega/c$, where

$$\left(2 - \frac{c^2}{c_s^2}\right)^2 - 4\sqrt{\left(1 - \frac{c^2}{c_s^2}\right)\left(1 - \frac{c^2}{c_p^2}\right)} = 0.$$

(It can be shown that this equation has a root in which $c_p > c_s > c$; it corresponds to the famous *Rayleigh wave* of seismometry.)

*4.17. According to Maxwell's equations, the static magnetic field $\mathbf{H} = (H_1, H_2, H_3)^{\mathrm{T}}$ in an inhomogeneous medium satisfies

$$\nabla \cdot (\mu \mathbf{H}) = 0, \quad \nabla \wedge \mathbf{H} = \mathbf{0},$$

where μ is a function of position. Suppose the interface $z = 0$ separates a medium in which $\mu = \mu_+ =$ constant from one in which $\mu = \mu_- =$ constant. By using Green's and Stokes' theorems show that a weak solution satisfies

$$[\mu H_3]_{z=0-}^{z=0+} = 0, \quad [H_1]_{z=0-}^{z=0+} = [H_2]_{z=0-}^{z=0+} = 0.$$

*4.18. Consider a time-varying electromagnetic field in a homogeneous conducting medium in which Ohm's law relates the current density and electric field by $\mathbf{j} = \sigma \mathbf{E}$, where σ is the conductivity, and suppose there is no charge density. Show that

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{\sigma}{\epsilon}\frac{\partial \mathbf{E}}{\partial t} = c^2 \nabla^2 \mathbf{E},$$

where $c^2 = 1/(\epsilon\mu)$. Hence show that, if $\mathbf{E} = \mathbf{E}(x, t)$, its components can be found in terms of solutions to the telegraph equation.

*4.19. As shown in Chapter 2, if the normal cone is $Q(\xi_0, \xi_1, \xi_2, \ldots) = 0$, where $\xi_i = \partial\varphi/\partial x_i$ are the components of the normal to the characteristics (wavefronts) $\varphi =$ constant, the ray cone is given by $\sum_i x_i \xi_i = 0$, where $x_i = \mu\, \partial Q/\partial \xi_i$. Take $x_0 = t$, $x_1 = x$ and $x_2 = y$ and show that, if the normal cone is

(i) $$a\xi_1 + b\xi_2 - \xi_0 = 0,$$

(ii) $$\frac{\xi_1^2}{a^2} + \frac{\xi_2^2}{b^2} - \xi_0^2 = 0,$$

(iii) $$\xi_1\xi_2 - \xi_0^2 = 0,$$

then the ray cone is

(i) $$\frac{x}{a} = \frac{y}{b} = -t,$$

(ii) $$a^2x^2 + b^2y^2 = t^2,$$

(iii) $$4xy = t^2.$$

Draw a typical cross-section ξ_0 = constant, t = constant in each case, including the case $a = b = \epsilon$, $\epsilon \to 0$ in (ii). Use the acquired knowledge to relate the local geometry near a point on a cross-section ξ_0 = constant of the normal cone to the geometry of a cross-section t = constant of the ray cone, and hence verify Fig. 4.11.

4.20. Suppose that

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial y^2}.$$

Show that, if $u(x,y) = v(p,q)$, where $p = \partial u/\partial x$ and $q = \partial u/\partial y$, then

$$\frac{\partial^2 v}{\partial p^2} = \frac{\partial^2 v}{\partial q^2}, \quad \text{as long as } \frac{\partial p}{\partial x} \neq \frac{\partial p}{\partial y}.$$

4.21. (i) Following (4.13), define $\nabla^2 \log r$ in two dimensions by the identity

$$\int_0^{2\pi}\int_0^\infty \psi \nabla^2(\log r)\, r\, \mathrm{d}r\, \mathrm{d}\theta = \int_0^{2\pi}\int_0^\infty (\log r)(\nabla^2\psi)\, r\, \mathrm{d}r\, \mathrm{d}\theta$$

in polar coordinates. By showing that

$$\int_0^{2\pi}\int_\epsilon^\infty \psi \nabla^2(\log r)\, r\, \mathrm{d}r\, \mathrm{d}\theta = \int_0^{2\pi}\left(\psi|_{r=\epsilon} + \epsilon\log\epsilon \left.\frac{\partial\psi}{\partial r}\right|_{r=\epsilon}\right)\mathrm{d}\theta$$
$$+ \int_0^{2\pi}\int_\epsilon^\infty (\log r)(\nabla^2\psi)\, r\, \mathrm{d}r\, \mathrm{d}\theta$$

(be careful about the direction of the normal and the orientation of the contour!) and taking the limit as $\epsilon \to 0$, show that

$$\nabla^2(\log r) = 2\pi\delta(x)\delta(y).$$

(ii) Show that, if $u = 0$ in $x < 0$ and $y < 0$ and

$$\frac{\partial^2 u}{\partial x\, \partial y} = \delta(x)\delta(y),$$

then $u = H(x)H(y)$. Does

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = \delta(x)\delta(y)$$

have a solution that vanishes in $x < 0$ and $y < 0$?

*4.22. The displacement $u(x,t)$ of an elastic string satisfies

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{for } -\infty < x < \infty.$$

Show that, if $u(x,0) = 0$ and $\partial u/\partial t\,(x,0)$ vanishes except for $|x| < \epsilon$, where it is a constant v_0, then $u = \epsilon v_0$ in $\epsilon - t < x < -\epsilon + t$, $t > \epsilon$. By taking a suitable limit show that, when $u(x,0) = 0$ and $\partial u/\partial t\,(x,0) = 2\delta(x)$, then $u(x,-t)$ is the Riemann function for the wave equation written in this form, in the domain $t < 0$ and with $(\xi,\eta) = (0,0)$. (This Riemann function is defined to satisfy

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 2\delta(x)\delta(t);$$

as in Exercise 4.2 the factor 2 is introduced because the equation is not in canonical form.) Show also that, if $u(x,0) = \delta(x)$ and $\partial u/\partial t\,(x,0) = 0$, then

$$u(x,t) = \frac{1}{2}\big(\delta(x+t) + \delta(x-t)\big).$$

5
Elliptic equations

In this chapter we will, as usual, begin by discussing some physical situations that are modelled by elliptic equations, as defined in a rather unfocused way in Chapter 3. Most of the examples involve scalar second-order equations, several of which are special cases, such as steady states, of evolution models discussed in Chapters 4 and 6.

The methods we will use in the subsequent analysis of these models are more *ad hoc* than those used on hyperbolic equations, for the simple reason that we have no general well-posedness statement analogous to that for the Cauchy problem for hyperbolic equations. Moreover, we will find that the influence of the data for elliptic problems, especially singularities in the boundary data, is much less localised and 'coherent' than it is for hyperbolic equations, where we recall that many kinds of singularities merely propagate along characteristics. These are the reasons why this chapter is longer than its predecessor; however, the most powerful tools we develop are Green's functions, which are direct analogues of the Riemann functions of Chapter 4, and eigenfunction expansions. Whereas most of the illustrations of Chapter 4 revolved around the wave equation, here the paradigm is Laplace's equation.

The range of applicability of elliptic equations is vast, as will be apparent from the following section where we will present models arising in gravitation, electromagnetism, mechanics, heat flow, chemical reactions and acoustics.

5.1 Models

5.1.1 Gravitation

Some of the most famous models in the history of applied mathematics lead to elliptic equations, the most revered perhaps being that of Newtonian gravitation. This is based on the observation that 'point' masses attract each other along the line joining them with a force proportional to the inverse square of the distance between them. Hence, with a suitable normalisation, the force field of a unit point mass (soon to be related to a Green's function) is the gradient of a potential

$$\phi = \frac{1}{r} \tag{5.1}$$

in spherical polar coordinates; ϕ is unique up to the addition of a constant. Summing over a distribution of such masses, the potential of a general density distribution $\rho(\mathbf{x})$ is

$$\phi(\mathbf{x}) = \iiint \rho(\mathbf{x}') \frac{d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \tag{5.2}$$

and it is a simple calculation to show that, away from any matter, ϕ satisfies the famous Laplace's equation

$$\nabla^2 \phi = 0. \tag{5.3}$$

Functions satisfying this equation are said to be ***harmonic functions***. However, it is less easy to show that, in the presence of matter, ϕ satisfies ***Poisson's equation*** in the form

$$\nabla^2 \phi = -4\pi\rho. \tag{5.4}$$

The most elementary procedure is to consider the integral in (5.2) with a small sphere around $\mathbf{x}$ excised, write down its Laplacian and use Green's theorem, but we will see this in a more general setting in §5.5.

Presented this way, the theory of gravitation does not seem to be a problem in differential equations, because the solution of Poisson's equation (5.4) is known to be (5.2). But often this formula is tedious to compute and we will soon see that the partial differential equation formulation is of inestimable conceptual value.

5.1.2 Electromagnetism

Almost equally venerable are models for electrostatics and magnetostatics, which can be derived at a glance from §4.7.2. In the absence of any charge ρ and current $\mathbf{j}$, and with $\partial/\partial t = 0$ in a steady state, we find

$$\nabla \wedge \mathbf{H} = \nabla \wedge \mathbf{E} = \mathbf{0}. \tag{5.5}$$

Hence both the magnetic and electric fields are the gradients of potentials, which, like (5.2), are only defined to within an additive constant; but they both satisfy Laplace's equation because, by (4.83) and (4.84), $\nabla \cdot \mathbf{H} = \nabla \cdot \mathbf{E} = 0$. Now we have a more clear-cut partial differential equation situation because these electromagnetic fields can be modelled as if they are created at boundaries, or at infinity, where $\mathbf{E}$ and $\mathbf{H}$ satisfy certain conditions. We will not give these conditions here except to say that the commonly occurring problem of finding the electric field outside a charged perfect conductor demands that the electric potential ϕ, where $\mathbf{E} = -\nabla\phi$, is constant on the conductor. The charge, which is confined to an exceedingly thin layer near the surface of the conductor, turns out to have a density proportional to the normal derivative $\partial\phi/\partial n$, a fact that can be inferred by reinstating the charge density ρ, so that $\nabla^2\phi = -\rho/\epsilon$, where ϵ is the permittivity, and using (5.2). The problem of designing a lightning conductor involves solving Laplace's equation with ϕ constant on the conductor and calculating where $\partial\phi/\partial n$ is largest, because this is where the lightning will strike.

Indeed, many everyday situations involve charge distributed on a long thin straight wire and it is easy to see that in such cases the electric potential in plane polar coordinates is proportional to $\log r$. Only the field, not the potential, goes to zero as we move away from the wire, in distinction to the potential (5.1) which decays as we move away from a point charge.

More commonly, we have to model currents flowing in electrically conducting materials. Then, if the current flow is steady, we can often use the experimentally observed *Ohm's law*

$$\mathbf{j} = \sigma \mathbf{E}, \tag{5.6}$$

where $\sigma > 0$ is the conductivity, taken to be constant for simplicity. When we take this in conjunction with Maxwell's equations (5.5), we have that $\mathbf{j} = -\sigma \nabla \phi$, where yet again ϕ satisfies $\nabla^2 \phi = 0$. Note that if we then needed to find the magnetic field we would have to solve

$$\nabla \wedge \mathbf{H} = \mathbf{j}, \quad \nabla \cdot \mathbf{H} = 0; \tag{5.7}$$

we can take the curl of the first equation to obtain a vector version of Poisson's equation, but (5.7) itself is an as-yet-unclassified system for $\mathbf{H}$ to which we will return in Chapter 9.

There are many, many other interesting situations involving Maxwell's equations and Ohm's law which we do not have space to mention in detail here. For example, it is often possible to reduce the model to ordinary differential equations when currents only flow in wires, and to two-dimensional partial differential equations for flow in metal sheets.

5.1.3 Heat transfer

We will use this even more familiar discipline to begin to write down some concrete problems for elliptic equations. Suppose heat is being conducted in a medium D whose temperature is $T(\mathbf{x})$ and in which there is a volumetric heat source $f(\mathbf{x})$, as in the element of an electric fire, or in food cooked in a microwave oven, or in the radioactive decay process in the core of the earth. Now Fourier's law of heat conduction states that the heat flux is given, analogously to (5.6), by

$$\mathbf{q} = -k \nabla T, \tag{5.8}$$

where $k > 0$ is the thermal conductivity. Hence conservation of energy requires that, for any region D in the material,

$$\iint_{\partial D} -k \frac{\partial T}{\partial n} \, \mathrm{d}S = -\iiint_D \nabla \cdot (k \nabla T) \, \mathrm{d}\mathbf{x} = \iiint_D f \, \mathrm{d}\mathbf{x} \tag{5.9}$$

in a steady state; here, as usual, $\partial/\partial n$ denotes the derivative along the outward normal to the boundary ∂D. This is true over any region in the conductor so that, when k is constant, we again retrieve Poisson's equation

$$k \nabla^2 T = -f(\mathbf{x}). \tag{5.10}$$

We are almost always given a boundary value problem for T, there being three very common situations.

The Dirichlet problem

Suppose the boundary is an excellent thermal conductor (e.g. a metal). Then its temperature is nearly constant, or at least a prescribed function of position on the boundary:

$$T = T_0(\mathbf{x}) \quad \text{on } \partial D. \tag{5.11}$$

The Neumann problem

Suppose the region outside D is a very bad conductor, such as air. Then the normal heat flux is zero or, more generally, a prescribed function of position:

$$-k\frac{\partial T}{\partial n} = q(\mathbf{x}) \quad \text{on } \partial D. \tag{5.12}$$

The Robin problem

The *Robin* problem arises when the heat flux at the boundary is proportional to the difference between the boundary temperature and some ambient temperature. This might be the case if D is a solid (such as your skin) around which a relatively hot (or cold, usually) fluid at temperature $T_0(\mathbf{x})$ flows, and heat transfer takes place across a thin 'boundary layer' in the fluid. There are then experimental and theoretical reasons to write down

$$-k\frac{\partial T}{\partial n} = h\,(T - T_0(\mathbf{x})) \quad \text{on } \partial D, \tag{5.13}$$

where h is called a ***heat transfer coefficient***. Thermodynamics requires that $h \geqslant 0$, so that heat flows from hotter regions to colder ones.

On physical grounds (5.12) is suspicious because, assuming D is bounded, we could not expect to be able both to prescribe the heat flux on ∂D and yet have a steady state in which T is independent of time. We can quantify this suspicion by integrating (5.10) over D and using Green's theorem to yield

$$\iiint_D 1 \cdot f(\mathbf{x})\,\mathrm{d}\mathbf{x} - \iint_{\partial D} 1 \cdot q\,\mathrm{d}S = 0 \tag{5.14}$$

for the Neumann problem. Here we have inserted the factors of unity to highlight that this is simply a statement of the Fredholm Alternative; the 'right-hand side' of the Neumann problem, which is a combination of f and q, must be 'orthogonal' to any relevant eigenfunctions, in this case constants. Moreover, when (5.14) is satisfied, even though T exists, it is undetermined to within the addition of any eigenfunction. We will see later that this non-existence/non-uniqueness does not happen with (5.13) if $h > 0$.

There are countless other variations and applications of the Laplace and Poisson equations, and we will cite only three very briefly. First, since Fick's law of molecular diffusion, namely that mass flux is proportional to concentration gradient, is mathematically identical to Fourier's law, the mathematics of heat transfer by conduction and of mass transfer by diffusion are identical.[53] Second, there are numerous situations where heat and mass transfer are intimately coupled, as in the case of many commonly occurring chemical reactions. We will study these in more detail in Chapter 6 but for now we simply note that the temperature in a steady-state reaction might be modelled by the uncoupled thermal model (5.10) in

[53] In fact, they are both based on the microscopic models of Brownian motion, which will be discussed further in a simple-minded way in §6.1.

regions where the source term is a function, possibly nonlinear, of T. This function represents local heat generation near the reacting regions and is positive or negative depending on whether the reaction is exothermic or endothermic, respectively. Finally, it is often important to consider convective and radiative heat transfer as well as conduction. The former is relatively easy to incorporate because we simply have to include a term $\rho c \iiint \mathbf{v} \cdot \nabla T$ in (5.9), where $\mathbf{v}$ is the velocity, ρ is the density and c the specific heat.[54] This leads to the *convection–diffusion* elliptic equation

$$\rho c \mathbf{v} \cdot \nabla T = k \nabla^2 T + f(\mathbf{x}), \tag{5.15}$$

where, if we are lucky, $\mathbf{v}$ is prescribed independently as the result of some uncoupled dynamics model. Alas, the dynamics and heat transfer are all too often coupled.

Radiation is much more difficult to model, as we shall see in Chapter 6, but a simple case is that in which a 'black body' radiates from its boundary, and all the heat transfer in the interior is by conduction alone. On experimental and theoretical grounds we can then write down the Stefan–Boltzmann law

$$q(\mathbf{x}) = \epsilon(T^4 - T_0^4) \tag{5.16}$$

in (5.12), where, again, T_0 is an ambient temperature (now measured in absolute terms) and ϵ is a constant called the *emissivity* (see [41] and §6.6.1.2); note that, when T is close to T_0, (5.16) can be approximated by (5.13) with $h = 4\epsilon T_0^3$.

All discussion of such nonlinear effects, be they from chemical reactions or radiation, is deferred to the end of the chapter.

5.1.4 Mechanics

5.1.4.1 Inviscid fluids

Continuum mechanics provides another justifiably famous source of elliptic equations. One of the most familiar is that of inviscid incompressible hydrodynamics [27]. The fact that the fluid is incompressible means that the velocity field $\mathbf{v}(\mathbf{x}, t)$ satisfies

$$\nabla \cdot \mathbf{v} = 0, \tag{5.17}$$

but it is a long story to prove that many inviscid flows can be well approximated by writing down the condition that the *vorticity* $\nabla \wedge \mathbf{v}$ satisfies

$$\nabla \wedge \mathbf{v} = \mathbf{0},$$

and thus that $\mathbf{v}$ is the gradient of a potential ϕ. We will discuss this point further in Chapter 9, but, for the moment, let us assume that ϕ exists, although it is only defined to within addition of an arbitrary function of time, and is harmonic from (5.17). Typical boundary conditions now take the form of homogeneous Neumann conditions because the relative velocity is tangential to any stationary impermeable

[54] However, if the material is a gas we must be much more careful about our definition of specific heat and also to incorporate the 'work done against compression'.

boundary;[55] this condition is associated with the fact that, for these inviscid flows, we only have Laplace's equation to solve, and hence in 'streaming flows' we have a Neumann problem in an infinite domain exterior to an obstacle. A more interesting model is that for a *vortex*, as may occur, approximately, in emptying the bath, where we could seek a two-dimensional flow in which the velocity is azimuthal, i.e. in the θ direction in plane polar coordinates (r, θ), getting faster and faster as we approach the 'core' of the vortex $r = 0$. An inspection of (5.3) shows that a suitable potential, indeed the only one, is proportional to θ. We also note that this flow provides a non-constant, albeit multiple-valued, solution of the Neumann problem with homogeneous boundary conditions in a circular annulus centred at the origin, and it gives an early warning that the connectivity of D may be important in considering the uniqueness of solutions of elliptic equations.

5.1.4.2 Membranes

The easiest way to visualise solutions of Laplace's equation is by looking at the shape of a membrane such as a drum, whose boundary is deformed slightly out of plane. Like a soap film, the membrane is assumed to be capable of supporting a tension T which is assumed to be isotropic, i.e. the same in all directions. Then a normal force balance reveals that the displacement w satisfies $\partial^2 w/\partial t^2 = a_0^2 \nabla^2 w$, where $a_0^2 = T/\rho$, ρ being the surface density of the membrane; this is because the components of the tension forces out of the plane add up to T multiplied by the small mean curvature, which is approximately $\nabla^2 w$. (In one dimension, this model reduces to that for waves on an elastic string, which is just the familiar wave equation (4.1).) Thus the equilibrium displacement is indeed a two-dimensional harmonic function with Dirichlet boundary conditions, assuming the perimeter is prescribed and nearly planar. Also, it is easy to see that the application of a pressure difference p across the membrane enables us to visualise solutions of Poisson's equation with p as the right-hand side.

5.1.4.3 Torsion

A less well-known, but no less important, model leading to the Laplace equation in two dimensions arises in the study of torsion in linear elasticity. Torsion bars are often used in motor car suspensions, and, in the simplest configuration, they are metal circular cylinders which are stress free except for equal and opposite torques about the axis which are applied at each end of the cylinder. To describe this we will again eschew the conventional tensorial formulation of elasticity theory and proceed on the justifiable assumption that the displacement from the unstressed state in any cross-section of the torsion bar aligned along the z direction is a rigid body rotation about the z axis which varies linearly in z (Fig. 5.1).

Hence we write the displacement as

$$\mathbf{u} = \alpha\,(-yz, xz, w(x, y))\,, \tag{5.18}$$

[55]This is in contrast to the paint flow of §1.1, where the paint adheres to the wall.

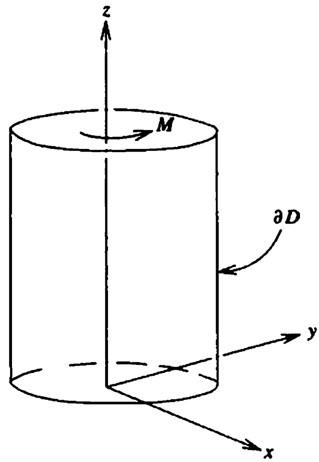


Fig. 5.1 Torsion of a bar.

where α is constant and represents the amount of twist applied to the bar.[56] Now this vector is divergence free, so that material is only undergoing shear and not compression or expansion. Thus, from (4.73) of §4.7, in equilibrium $\nabla\wedge(\nabla\wedge\mathbf{u}) = \mathbf{0}$, so that

$$\nabla^2 w = 0. \tag{5.19}$$

The key modelling feature of any torsion bar is that the shear forces on the curved boundary ∂D of the cross-section of the bar are zero. Those in the x and y directions are automatically zero because the displacement in the section is a rigid body one, but the shear force axially (in the z direction) must also vanish. It can be shown, using the ideas at the beginning of §4.7, that the displacement in this direction contributes

$$\mu\alpha\left(\frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, 0\right)\cdot\mathbf{n}$$

to this shear, where $\mathbf{n}$ is the outward normal to ∂D. Meanwhile, the in-plane components of $\mathbf{u}$ contribute

$$\mu\alpha\left(\frac{\partial}{\partial z}(-yz), \frac{\partial}{\partial z}(xz), 0\right)\cdot\mathbf{n}$$

and in total we are left with the Neumann condition

$$\frac{\partial w}{\partial n} = (y, -x, 0)\cdot\mathbf{n} \quad \text{on } \partial D. \tag{5.20}$$

Mercifully, but not surprisingly, the orthogonality condition corresponding to (5.14) is satisfied automatically.

[56]Notice the apparently counterintuitive fact that the axial displacement (in the z direction) is independent of z.

Finally, we must specify the tractions, i.e. the stresses that are applied to the ends of the bar to produce the torsion. The shear stresses on these ends comprise the two terms

$$\mu\alpha(-y,x,0)+\mu\alpha\left(\frac{\partial w}{\partial x},\frac{\partial w}{\partial y},0\right),$$

and hence the total moment applied about the origin in any cross-section is

$$M=\mu\alpha\iint_D\left(x^2+y^2-y\frac{\partial w}{\partial x}+x\frac{\partial w}{\partial y}\right)\mathrm{d}x\,\mathrm{d}y. \tag{5.21}$$

The fundamental problem of torsion is to solve (5.19) subject to (5.20) and (5.21) in order to relate M to α and hence determine the 'torsional rigidity'. But many torsion bars are hollow and we must remember our earlier warning about non-uniqueness in multiply-connected domains.

5.1.4.4 *Plane strain and slow viscous flow*

Continuing the previous discussion, suppose we have an infinitely long bar which is not subject to torsion but whose curved boundary does have tractions applied to it perpendicular to the axis. We now have what is called ***plane strain***[57] with

$$\mathbf{u}=(u(x,y),v(x,y),0), \tag{5.22}$$

which implies there are both shear and tensile/compressive stresses in the bar. We now need both terms in the right-hand side of the equations of linear elasticity (4.73), so that

$$\mu\nabla^2u+(\lambda+\mu)\left(\frac{\partial^2u}{\partial x^2}+\frac{\partial^2v}{\partial x\,\partial y}\right)=0,$$
$$\mu\nabla^2v+(\lambda+\mu)\left(\frac{\partial^2u}{\partial x\,\partial y}+\frac{\partial^2v}{\partial y^2}\right)=0.$$

It is a simple matter to cross-differentiate and show that both u and v satisfy the ***biharmonic equation***[58]

$$\nabla^4u=\nabla^4v=0, \tag{5.23}$$

where

$$\nabla^4=(\nabla^2)^2=\frac{\partial^4}{\partial x^4}+2\frac{\partial^4}{\partial x^2\partial y^2}+\frac{\partial^4}{\partial y^4}.$$

In practice, it turns out often to be convenient to reformulate this model in terms of the stresses (see Exercise 5.1). This is very important to gain insight into

[57]The displacements are non-zero only in the (x,y) plane. The complementary situation in which the displacement has the form $(0,0,w(x,y))$ is called *antiplane strain*, and it is easy to see that $\nabla^2w=0$.

[58]Although the biharmonic equation is fourth order, its solution turns out to be so similar to that of an elliptic second-order equation that we describe it as elliptic; when it is written as a first-order system, it is indeed elliptic according to the ideas of §§3.3 and 3.4.

the kinds of boundary conditions that might be imposed on the biharmonic equation, and we note that the application of tractions on ∂D involves the prescription of a two-dimensional vector there. This vector is in fact a linear combination of the first derivatives of u and v, and hence we must be prepared for 'twice as much' boundary data as for Laplace's equation.

Two-dimensional problems of slow viscous flow also lead to the biharmonic equation. There the assumption of a shear viscosity μ in a fully two-dimensional flow[59] gives that the forces generated by the pressure p must balance the shear forces, in the absence of inertia. Hence, with velocity $\mathbf{v} = (u, v)^{\mathrm{T}}$, we can follow the argument leading to (4.73), with $\lambda = 0$, to derive the model

$$\nabla p = \mu \nabla^2 \mathbf{v} = \mu \begin{pmatrix} \nabla^2 u \\ \nabla^2 v \end{pmatrix} \tag{5.24}$$

and, assuming incompressibility as in (5.17),

$$\nabla \cdot \mathbf{v} = 0. \tag{5.25}$$

Now, following the idea of introducing a potential, as already used several times in this chapter, we eliminate (5.25) by writing $u = \partial\psi/\partial y$ and $v = -\partial\psi/\partial x$, where ψ is a ***stream function***. Then we can take the curl of (5.24) to show that u, v and ψ all satisfy the biharmonic equation (p only satisfies Laplace's equation, but, regrettably, in a realistic model boundary conditions can never be prescribed just for p). Moreover, because the flow is viscous we expect that on an impermeable boundary $\mathbf{v}$ itself is prescribed rather than just $\mathbf{v}\cdot\mathbf{n}$, as was the case for an inviscid fluid; hence we again have two boundary conditions to be satisfied.

We note that, if a slow viscous flow occurs in a porous medium such as a sugar-lump, sponge or rock, the simplest model is to assume *Darcy's law* to relate the velocity and pressure. This means that the slow viscous flow model (5.24) is abandoned and replaced by Darcy's experimental observation that, on a scale much bigger than the pore size,

$$\mathbf{v} = -\frac{k}{\mu}\nabla p \tag{5.26}$$

in the absence of gravity, which is clearly analogous to the Fourier, Fick and Ohm laws mentioned above; k is a positive constant called the permeability and it depends only on the geometry of the pores. However, there is a serious question about what we mean by $\mathbf{v}$ in such a complicated configuration and we must be careful to remember that the fluid does not occupy the whole of space. Nonetheless, by regarding $\mathbf{v}$ as measuring the mass flux, it can still be argued that $\nabla\cdot\mathbf{v} = 0$ for an incompressible fluid, and we end up with Laplace's equation again. It is uncommon for an increase in the geometric complexity of a problem to be accompanied by a decrease in its analytical complexity!

Not surprisingly, these examples enable much 'technology transfer' between fluid and solid mechanics.

[59] Rather than the quasi-one-dimensional flows such as the paint flow of §1.1.

5.1.5 Acoustics

There is no better illustration of the delicacy of elliptic differential equations than the vitally important science of acoustics, i.e. small-amplitude wave propagation in fluids or solids.[60] We will only consider the former here, and then only in the so-called *frequency domain* where we assume the sound field is being generated by a loudspeaker operating at just one frequency ω. Then, in the simplest case, when we write $\phi = \Re\,(u(\mathrm{x})e^{-i\omega t})$ in the wave equation $\partial^2\phi/\partial t^2 = a_0^2\nabla^2\phi$, we find

$$(\nabla^2 + k^2)u = 0, \tag{5.27}$$

where $k = \omega/a_0$; (5.27) is known as *Helmholtz' equation*.[61] The seemingly similar model

$$(\nabla^2 - k^2)u = 0, \tag{5.28}$$

called the *modified Helmholtz equation*, has absolutely nothing to do with acoustics, as will soon become evident from a study of its mathematical properties. However, (5.27) and (5.28) can be related in the context of chemical reactions because we have seen that, if the heat source is proportional to the temperature, (5.10) reduces to (5.27) or (5.28) when the reaction is exothermic or endothermic, respectively.

There is also another clear physical reason to expect the properties of solutions to (5.27) and (5.28) to differ widely. Suppose we have any of the boundary conditions (5.11)–(5.13) but with zero forcing. With these homogeneous boundary conditions, $u = 0$ is a solution of the problem, but is it the only one? We would expect an acoustic resonator, modelled by (5.27), to be able to vibrate of its own accord; in other words, we would expect the existence of eigenfunctions, and probably many of them. We will soon see that a trivial application of Green's theorem shows that, whereas (5.28) admits only the zero solution, there is no such uniqueness result for (5.27). Of course, our experience suggests that non-zero solutions only exist at certain 'eigenfrequencies' $\omega = \omega_i$, but this is not always true. Our investigation into this problem will inevitably involve a little of the vast theory of eigenfunctions of elliptic differential operators, but only the rudiments are needed later in this chapter.

One other crucial remark needs to be made concerning acoustics in the frequency domain, and that concerns the model in cases where the physical domain of interest extends to infinity. Since the model describes wave motion, we expect it to be able to distinguish between situations in which waves are 'incoming' from those in which they are 'outgoing'; this suggests that we need more precise conditions at infinity for Helmholtz' equation than we would for Laplace's equation in an infinite physical domain.

[60] In solids we usually obtain vector elliptic equations, which are closely analogous to Maxwell's equations in the frequency domain.

[61] The complex time dependence of ϕ means that we need to take complex combinations of real solutions of Helmholtz' equation. The choice of $e^{-i\omega t}$ rather than $e^{i\omega t}$ is made in order to simplify later formulæ.

5.1.6 Aerofoil theory and fracture

We conclude this section with two examples which highlight the global importance of the specification of any local singular behaviour that may occur in the solution of elliptic equations. A trivial illustration of the sort of thing that can happen is the Dirichlet problem for Laplace's equation

$$\nabla^2 u = 0$$

in the open sector

$$0 < r < 1, \quad 0 < \theta < \frac{\pi}{2}$$

in plane polar coordinates, with $u = 0$ on the boundary. If we insist that u is bounded at the corners of the domain, the only solution is $u = 0$. However, it is easy to verify that

$$u \propto \left(r^2 - \frac{1}{r^2}\right) \sin 2\theta$$

satisfies all the conditions except for boundedness at the origin. Thus a singularity at one point of the boundary makes its presence felt *everywhere*. Such singularities could occur in any of the aforementioned situations, but we just mention two striking examples which perhaps have more everyday technological importance than any others in this book.

5.1.6.1 Aerofoils

In §5.1.4.1 above, we blandly asserted that the boundary condition on, say, an impermeable obstacle placed in an irrotational streaming flow, such as in a wind tunnel, is the Neumann condition for a velocity potential satisfying Laplace's equation. Moreover, it is well known that a properly designed aerofoil has a sharp trailing edge, i.e. ∂D must *not* be smooth if the aerofoil is to work well. The details of the fluid mechanics of this situation are too intricate to enter into here [29] but one simple scenario can be plausibly laid out. Suppose that a two-dimensional aerofoil with surfaces $y = f_\pm(x)$, $0 < x < c$, is so thin and so nearly aligned with the flow that it is always close to the line $y = 0$, $0 < x < c$, and that the free stream velocity in the x direction is U (see Fig. 5.2). Then, as in the linearised

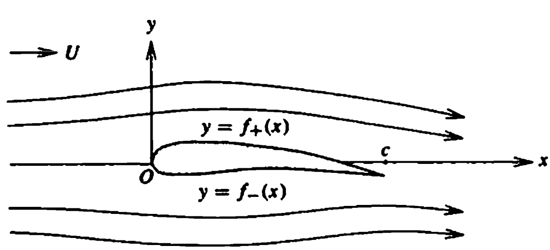


Fig. 5.2 Flow past an aerofoil.

models (3.11) and (4.70), ϕ is almost equal to Ux and so we can write $\phi = Ux + \tilde{\phi}$ to obtain

$$\nabla^2\tilde{\phi} = 0, \quad |\nabla\tilde{\phi}| \to 0 \quad \text{as } |\mathbf{x}| \to \infty, \tag{5.29}$$

but we will see later that we must consider the decay rate at infinity more carefully than we have in (5.29) if we are to explain the theory of flight. Lastly, we can replace the exact condition of tangential flow, namely $\partial\phi/\partial y = f'_{\pm}(x)\,\partial\phi/\partial x$ on the aerofoil $y = f_{\pm}(x)$, by

$$\frac{\partial\tilde{\phi}}{\partial y} = Uf'_{\pm}(x) \quad \text{on } y = \pm 0,\, 0 < x < c. \tag{5.30}$$

We are thus led to a boundary value problem for a harmonic function in a multiply-connected domain; we might regard the discontinuity that inevitably occurs across the segment $0 < x < c$ of the x axis as a line of singularities, and in particular we may expect especially violent singularities at the 'leading' and 'trailing' edges $(0,0)$ and $(0,c)$. One problem for elliptic theory is to decide what to do about these singularities, which will also be crucial to our understanding of the theory of flight.

5.1.6.2 Brittle fracture

Our final example concerns the modelling of cracks in perfectly elastic solids. A realistic configuration is illustrated in Fig. 5.3(a) for a crack $y = 0$, $-c < x < c$, in a material in plane strain under tension in the y direction. However, we will confine ourselves to the easier cracks that can occur in antiplane strain with a displacement field $\mathbf{u} = (0, 0, w(x,y))$ where w is a harmonic function. There is no traction on the crack surfaces, so that

$$\frac{\partial w}{\partial y} \to 0 \quad \text{as } y \to 0,\, -c < x < c \tag{5.31}$$

from above or below, so that the problem can still be represented as in Fig. 5.3(a). Also, assuming symmetry,

$$w = 0 \quad \text{on } y = 0,\, |x| > c, \tag{5.32}$$

where the material is pristine. Finally, we suppose the crack is subject to a uniform shearing at large distances, modelled by

$$w \to \tau y \quad \text{as } |y| \to \infty, \tag{5.33}$$

where τ is a constant.

The symmetry about $y = 0$ means that we need only solve the problem in a half-plane, in contrast to the aerofoil model (5.30), where we need to find a harmonic function in the doubly-connected region exterior to the aerofoil (unless, of course, the aerofoil is symmetric and $f_+ = -f_-$).

The boundary conditions (5.31) and (5.32) are an example of what is called a ***mixed boundary problem***, the data being part Neumann and part Dirichlet. This

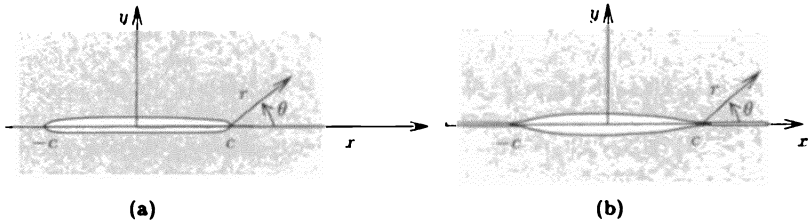


Fig. 5.3 (a) Brittle fracture in pristine material; (b) partly opened pre-existing crack.

switch engenders possible singularities at the tips of the crack which we can readily see by separating the variables in local polar coordinates (r, θ), say near $x = c$, $y = 0$. The function $r^\lambda \sin \lambda\theta$ is a local 'eigensolution' as long as

$$\lambda = n + \frac{1}{2},$$

where n is an integer, because it is only for these values of λ that $w = 0$ on $\theta = 0$ and $\partial w/\partial\theta = 0$ on $\theta = \pi$. As in aerofoil theory, we shall see that this 'singular behaviour' has far-reaching implications for the predictions of the mathematical models. Even at this early stage one might, for example, enquire as to the difference between models for a crack and for a 'contact' region. The latter could arise if the two slabs in Fig. 5.3(b), each containing a slight indentation in $-c < x < c$, had been brought together; how would the mathematics distinguish this situation from that in Fig. 5.3(a), which involves a crack surrounded by pristine material?

With all these examples as motivation, we now embark on as general an account of the theory of elliptic partial differential equations as we are able to give in a book of this type. We will clearly have much to say about Laplace's equation and the effects of singularities in the data, and we will confine most of our comments to two-dimensional problems. However, a great deal of our discussion generalises quite easily to higher dimensions and sometimes to higher-order and vector equations as well.

5.2 Well-posed boundary data

5.2.1 The Laplace and Poisson equations

We could begin our discussion by repeating the procedures of Chapter 4 and writing Laplace's equation in characteristic variables $z = x + iy$ and $\bar{z} = x - iy$ as

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 4\frac{\partial^2 u}{\partial z\,\partial\bar{z}} = 0,$$

noting that u is the sum of a function of z and a function of $\bar{z}$. Then the solutions satisfying analytic Cauchy data, say, $u(x, 0)$ and $\partial u/\partial y\,(x, 0)$ given, can be written down from d'Alembert's formula (4.38) as

$$u(x,y) = \frac{1}{2}\big(u(z,0) + u(\bar{z},0)\big) + \frac{1}{2\mathrm{i}}\int_{\bar{z}}^{z} \frac{\partial u}{\partial y}(\xi,0)\,\mathrm{d}\xi, \tag{5.34}$$

but we already know from our discussion of the Cauchy–Riemann system and analytic continuation that this recipe is ill-posed. This is a relief because we can see from the models in §5.1 that it would be a disaster if well-posedness for Laplace's equation demanded analytic Cauchy data.

Let us therefore reconsider Poisson's equation in the form

$$\nabla^2 u = f(x,y). \tag{5.35}$$

Motivated by the models in §5.1, we begin by considering (5.35) in a closed bounded region D but with only a single boundary condition rather than the Cauchy data needed for (5.34). All possible linear conditions are encompassed by the Robin condition, which we will write in the form

$$\alpha u + \beta\frac{\partial u}{\partial n} = g \quad \text{on } \partial D, \tag{5.36}$$

where α, β and g are given functions of position, and we recall that $\partial/\partial n$ denotes the outward normal derivative; we expect α and β to be of the same sign from (5.13). Our immediate worry that we may have allowed the solution too much freedom by reducing the number of boundary conditions is unfounded because of the following theorem.

Uniqueness theorem *Suppose u satisfies $\nabla^2 u = f$ in a bounded domain D with smooth boundary ∂D, with $\alpha u + \beta\,\partial u/\partial n = g$ on ∂D. Then, if a solution exists, it is unique provided that α never vanishes and $\beta/\alpha \geqslant 0$.*

An elementary proof uses Green's theorem over D in the form

$$\iint_D \nabla\cdot(U\nabla U)\,\mathrm{d}x\,\mathrm{d}y = \int_{\partial D} U\frac{\partial U}{\partial n}\,\mathrm{d}s, \tag{5.37}$$

where U is the difference between any two solutions of (5.35) and (5.36). Since α never vanishes,

$$\iint_D |\nabla U|^2\,\mathrm{d}x\,\mathrm{d}y = -\int_{\partial D} \frac{\beta}{\alpha}\left(\frac{\partial U}{\partial n}\right)^2\mathrm{d}s; \tag{5.38}$$

considering the signs of the two sides of this equation, we see that both sides must vanish and thus $|\nabla U| = 0$ almost everywhere in D. This means that U is constant almost everywhere and hence zero from the boundary condition.[62] Although we have stated it in two dimensions, the theorem holds in any number of dimensions, with the same proof.

[62]We cannot exclude the possibility that U is non-zero on a set of measure zero, but then U would probably not satisfy Laplace's equation on such a set and we will not deal with such 'pathologies' in this book.

As a corollary, we note that, when $\alpha = 0$,

$$\iint_D |\nabla U|^2 \,\mathrm{d}x\,\mathrm{d}y = 0,$$

so that u (if it exists) is non-unique to within addition of a constant. Thus, for example, the solution to the torsion problem of §5.1.4.3 is only determined to within an additive constant in w, which just corresponds to moving the bar bodily along its axis. It is also possible to give specific examples to show that, if $\beta \neq 0$ and α/β changes sign on ∂D, or if $\alpha/\beta < 0$, uniqueness can no longer be assured (see Exercise 5.4). We recall that the latter case corresponds to heat flow from cold to hot in (5.13). Also, it is clear from Green's theorem that, when $\alpha = 0$, no solution exists unless, as in (5.14),

$$\iint_D f \,\mathrm{d}x\,\mathrm{d}y = \int_{\partial D} \frac{g}{\beta}\,\mathrm{d}s. \tag{5.39}$$

The uniqueness theorem and its corollary have been stated assuming that ∂D and the data are as smooth as necessary. We must also remember the example at the beginning of §5.1.6, where, even though Dirichlet data were prescribed at all points of ∂D except one, the existence of a singularity at that point resulted in a solution that was non-zero in the interior of D. This can happen whether or not ∂D is smooth, and it will cause us much trouble later in this chapter.

These results will come as no surprise to experts in two-point boundary value problems with Robin boundary conditions for ordinary differential equations, but here, in addition to the singularity problem we have mentioned, we have to worry about the possibility that D is not simply connected. Suppose, for example, we adopt the following commonly used device to solve the torsion problem (5.19)–(5.21). There is often a computational and theoretical advantage in working with a Dirichlet problem rather than the Neumann problem for w, so we introduce its harmonic conjugate ψ for which a simple calculation reveals the Dirichlet problem

$$\nabla^2\psi = 0,$$

where the zero-stress condition (5.20) requires that

$$\psi - \frac{1}{2}(x^2 + y^2) = \text{constant} \quad \text{on } \partial D,$$

the constant being arbitrary when D is simply connected. Things are even easier computationally if we can take the constant to be zero and write[63] $\psi = \frac{1}{2}\left(x^2 + y^2\right) + \widetilde{\psi}$, so that

$$\nabla^2\widetilde{\psi} = -2 \qquad \text{with} \qquad \widetilde{\psi} = 0 \quad \text{on } \partial D. \tag{5.40}$$

Now, if D is an annulus, as is often the case in practice, $\widetilde{\psi}$ is a different constant on each component of the boundary of the annulus, but what is the difference

[63] $\widetilde{\psi}$ is called the *Prandtl stress function*.

between these constants? We certainly do not have a well-posed Dirichlet problem without this knowledge. The answer lies in the fact that, in transforming from the physical variable w to a stress function ψ, we have introduced an indeterminacy and we must be sure that when we eventually retrieve w from ψ, by solving the Cauchy–Riemann equations, we have a physically acceptable displacement. In this case the w we retrieve is multi-valued unless

$$\int_{\partial D} \frac{\partial w}{\partial s}\, \mathrm{d}s = -\int_{\partial D} \frac{\partial \psi}{\partial n}\, \mathrm{d}s$$

is zero around any circuit in the annulus that cannot be shrunk to zero. Hence $\tilde{\psi}$ is uniquely defined by the condition that

$$\int_{\partial D} \frac{\partial \tilde{\psi}}{\partial n}\, \mathrm{d}s + \int_{\partial D} \frac{\partial}{\partial n}\left(\frac{1}{2}\left(x^2+y^2\right)\right) \mathrm{d}s = 0 \tag{5.41}$$

taken around any such circuit (the first term on the left-hand side is in any case a constant independent of the circuit).

This calculation reveals a common bugbear of working with 'potentials' such as ψ or $\tilde{\psi}$: they are often helpful theoretically and computationally but their arbitrariness (or so-called 'gauge invariance') calls for constant vigilance. This example, and the vortex problem in §5.1.4.1, illustrate the message that great care must always be taken when considering uniqueness for elliptic problems in multiply-connected domains.

5.2.2 More general elliptic equations

We have already seen in our discussion of the Helmholtz equation in the introduction that questions of existence and uniqueness can depend very sensitively on the coefficients of elliptic equations. Hence we will not enumerate statements about equations other than the Laplace and Poisson equations, except to say that the ideas above can be tried but they may or may not work. As an example, we cite the biharmonic equation $\nabla^4 u = 0$ with 'Dirichlet data' in which u and $\partial u/\partial n$ are prescribed on a closed boundary ∂D. The problem of uniqueness can be resolved by considering the difference $w = u - v$, for which

$$\begin{aligned} 0 &= \iint_D w\nabla^4 w\, \mathrm{d}x\, \mathrm{d}y = \iint_D \left(\nabla\cdot\left(w\nabla(\nabla^2 w)\right) - \nabla w\cdot\nabla(\nabla^2 w)\right) \mathrm{d}x\, \mathrm{d}y \\ &= -\iint_D \left(\nabla\cdot(\nabla^2 w\nabla w) - (\nabla^2 w)^2\right) \mathrm{d}x\, \mathrm{d}y = \iint_D (\nabla^2 w)^2\, \mathrm{d}x\, \mathrm{d}y, \end{aligned}$$

each divergence integrating to zero by Green's theorem. Hence $\nabla^2 w = 0$ and, because of the boundary data, $w = 0$.

The use of Green's theorem to prove uniqueness is only at all easy when the differential equation is in divergence form. This makes uniqueness theory difficult for vector equations, and for higher-order equations other than the biharmonic equation. Also, problems of mixed elliptic and hyperbolic type pose their own peculiar challenges for existence and uniqueness.[64] Fortunately, there is one other tool that we can use in many practical situations.

5.3 The maximum principle

It is almost obvious that, if Laplace's equation holds in a closed domain D, the maxima and minima of u occur on ∂D; one would never see a membrane, as modelled in §5.1.4.2, with a 'bump' in it, unless, of course, a pressure difference was applied across it. This is certainly true if we exclude points at which $\partial^2 u/\partial x^2$ and $\partial^2 u/\partial y^2$ both vanish because, if one is positive and the other negative in the interior of D, at any critical point where $\partial u/\partial x = \partial u/\partial y = 0$, we have

$$\frac{\partial^2 u}{\partial x^2}\frac{\partial^2 u}{\partial y^2} < 0 \leqslant \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2,$$

and so the critical point must be a saddle. To make this argument rigorous we note that, if an auxiliary function V satisfies Poisson's equation (5.35) with $f > 0$, then certainly one of $\partial^2 V/\partial x^2$ or $\partial^2 V/\partial y^2$ is positive and so V cannot have a maximum in the interior of D. It may have a minimum or saddle in the interior, but any maximum is on ∂D. Hence, if $\nabla^2 u = 0$ and we write $V = u + \epsilon r^2/4$ (where $\epsilon > 0$, $r^2 = x^2 + y^2$ and $(0,0)$ is assumed to be in D), $\nabla^2 V = \epsilon > 0$ and so any maximum of V is on ∂D. Hence

$$u + \frac{\epsilon r^2}{4} \leqslant M_V < M_u + \frac{\epsilon R^2}{4},$$

where M_V and M_u are the maximum values of V and u, respectively, on ∂D and R is the largest distance from the origin to ∂D. Letting $\epsilon \to 0$ in $u < M_u + \epsilon(R^2 - r^2)/4$, we see that any maximum of u also occurs on ∂D (and, similarly, any minimum).

The maximum principle enables us to rederive part of the uniqueness theorem in §5.2 and also to do much more, as we now see.

Uniqueness

For Dirichlet data for Poisson's equation, the difference between two solutions is a harmonic function vanishing on the boundary. The fact that it attains its maximum and minimum there means it is zero.

Comparison theorem

Consider two Poisson equations $\nabla^2 u_i = f(x, y)$, $i = 1, 2$, with the same right-hand side f and smooth Dirichlet data $g_i = g_i(x, y)$, such that $g_1 < g_2$ pointwise. Then $u_1(x, y) < u_2(x, y)$ in D because $u_2 - u_1$ attains its positive minimum on ∂D. Similarly, if $g_1 = g_2$ but the Poisson equations have different right-hand sides f_1 and f_2 with $f_1 > f_2$ pointwise, then $u_1(x, y) < u_2(x, y)$.

[64]Note that our accumulated knowledge of hyperbolic and elliptic equations can be used to help to understand mixed equations such as (3.13) which are elliptic and hyperbolic on either side of a 'sonic line', in the case that data is given on a curve Γ that crosses this line. Clearly we would have an overdetermined problem if we prescribed Cauchy data on the 'hyperbolic part' of Γ if that data enabled us to find both u and its normal derivative on the sonic line.

Continuous dependence on the data

This is a trivial extension of the comparison theorem when we let $g_1 \to g_2$ (from above and below). A similar result describes the dependence of u on f.

This last result is the first real evidence that the Dirichlet problem for Poisson's equation is in fact well posed. But we must remember that we have not yet proved existence!

In this book we will not delve much further into the vast number of uses to which the maximum principle and its generalisations can be put, although we note that the principle can also be used when the dependent variable u satisfies suitable differential inequalities, as we will see in §5.11, or more general elliptic equations. The question of existence is the hardest of all and requires a lengthy chain of arguments, of which the starting point is often the following methodology.

5.4 Variational principles

In certain circumstances, which occur almost as frequently as those under which the maximum principle applies, it may be possible that the solutions of elliptic equations are minima or other stationary points of a variational integral. Indeed, in subjects such as thermodynamics and nonlinear elasticity, it is common to model processes in terms of such minimisation procedures and only consider the associated Euler–Lagrange equations as partial differential equation problems *a posteriori*. This strategy has great advantages in the computer age, when discrete approximations to an integral (as, for example, in the finite element method) are very easy and convenient compared to discrete approximations of a differential equation. However, in this book we always take the viewpoint that the partial differential equation is the fundamental model.

As an example, supposing f is given and u minimises

$$E(u) = \iint_D \left(\frac{1}{2}|\nabla u|^2 + f(x,y)u\right) \mathrm{d}x\,\mathrm{d}y, \tag{5.42}$$

with or without boundary conditions, we would vary u to find

$$\begin{aligned} E(u+\eta) - E(u) &= \iint_D (\nabla\eta \cdot \nabla u + \eta f)\,\mathrm{d}x\,\mathrm{d}y + O(\eta^2) \\ &= -\iint_D \eta(\nabla^2 u - f)\,\mathrm{d}x\,\mathrm{d}y + O(\eta^2), \end{aligned}$$

neglecting boundary conditions. Because η is arbitrary, E can only attain a stationary value when u is a solution of Poisson's equation (5.35).

The advantages of working with $\min_u E(u)$ are not confined to numerical algorithms. For example, we could consider using any one of a number of optimisation algorithms analytically to construct sequences u_{n+1} such that $E(u_{n+1}) \leqslant E(u_n)$. Then, if we could prove that $\{u_n\}$, or at least a subsequence thereof, converges, the limit would, in many circumstances, be a 'weak solution' of Poisson's equation, i.e. a function which could be proved to satisfy Poisson's equation when multiplied by an arbitrary test function and integrated by parts. If we could prove enough about its regularity, we might then even be able to prove that this weak solution was a classical one in that the left- and right-hand sides of Poisson's equation were equal everywhere in D. This is not so unlikely because we know from Chapter 3 that elliptic equations cannot 'propagate' singularities; all the singularities we have encountered make their 'presence' felt everywhere, but their effect away from the boundary has always been very smooth, in fact analytic. Further details of this procedure can be found in [12], where iterations other than those suggested by $E(u)$ are used as starting points.[65]

We remark that variational principles have one of their most important practical applications in the study of eigenvalue problems for elliptic equations, as we shall see briefly in §5.7.1.

5.5 Green's functions

Green's functions provide the most important technique for gaining insight into the structure of solutions of linear elliptic equations. These functions are the analogues of Riemann functions for hyperbolic equations, but whereas Riemann functions are multidimensional generalisations of Green's functions for initial value problems for second-order ordinary differential equations, Green's functions for elliptic equations extend the theory of two-point boundary value problems for ordinary differential equations.

As usual, we begin with Poisson equations and, as in §4.2, we can proceed in two ways, both of which reach the same conclusion. Either we can go through a fairly lengthy analysis using classical functions, or we can take a short cut, requiring more 'infrastructure', by using generalised functions.

5.5.1 The classical formulation

We begin with the Dirichlet problem

$$\nabla^2 u = f \quad \text{in } D, \qquad u = g \quad \text{on } \partial D, \tag{5.43}$$

where D is a smooth, bounded, simply-connected domain in $\mathbb{R}^2$, and we write $\mathbf{x} = (x, y)$.

Our starting point is to recall that the two-point boundary value problem for the ordinary differential equation

$$\mathcal{L}u = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = f(x), \quad u(0) = u(1) = 0, \tag{5.44}$$

is formally solved by

$$u(\xi) = \int_0^1 f(x) G(x, \xi)\,\mathrm{d}x, \tag{5.45}$$

where

[65] A common one is that resulting from time-stepping in an evolution equation of the type considered in Chapter 6.

$$\mathcal{L}G = \frac{\mathrm{d}^2 G}{\mathrm{d}x^2} = 0 \quad \text{for } x \neq \xi, \tag{5.46}$$

with $G(0,\xi) = G(1,\xi) = 0$, G continuous and

$$\frac{\mathrm{d}G}{\mathrm{d}x}(\xi + 0, \xi) - \frac{\mathrm{d}G}{\mathrm{d}x}(\xi - 0, \xi) = 1. \tag{5.47}$$

Equation (5.45) states that u is the integral of f over $(0,1)$ weighted by a Green's function G which has a slope discontinuity at $x = \xi$ of just the right size to ensure that

$$\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} \int_0^1 f(x) G(x,\xi)\,\mathrm{d}x = f(\xi).$$

We can see that this is the case by simply multiplying (5.44) by G and (5.46) by u, subtracting, and integrating over $0 \leqslant x \leqslant 1$.[66]

Motivated by this result for ordinary differential equations, we seek a function $G(\mathbf{x}, \boldsymbol{\xi})$ such that, for each $\boldsymbol{\xi} \in D$,

$$\nabla^2 G = 0 \quad \text{for } \mathbf{x} \neq \boldsymbol{\xi}. \tag{5.48}$$

Now, as above, we multiply this equation by u and Poisson's equation (5.43) by G, subtract, and integrate over D to try to 'pick out' $u(\boldsymbol{\xi})$. To do this we need G to satisfy the following conditions.

1. We must have $G = 0$ on ∂D, or else we would be left with unwanted boundary contributions involving $\partial u / \partial n|_{\partial D}$.
2. We need G to have a suitable singularity at $\mathbf{x} = \boldsymbol{\xi}$ analogous to the 'kink' in (5.47).

The second condition is difficult to motivate, but if we guess that this singularity is isotropic, i.e. the behaviour of G is independent of the direction[67] of $\mathbf{x} - \boldsymbol{\xi}$, we are forced to try

$$G(\mathbf{x}, \boldsymbol{\xi}) = \text{constant} \cdot \log|\mathbf{x} - \boldsymbol{\xi}| + O(1) \quad \text{as } \mathbf{x} \to \boldsymbol{\xi}. \tag{5.49}$$

With hindsight, we choose the constant to be $1/2\pi$ and now we apply Green's theorem in the form

$$\int_{\partial(D - D_\epsilon)} \left(u \frac{\partial G}{\partial n} - G \frac{\partial u}{\partial n} \right) \mathrm{d}s = \iint_{D - D_\epsilon} \left(u \nabla^2 G - G \nabla^2 u \right) \mathrm{d}\mathbf{x}, \tag{5.50}$$

[66]Precisely the same argument applies when

$$\mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + p(x) \frac{\mathrm{d}}{\mathrm{d}x} + q(x),$$

as long as we replace $\mathcal{L}G$ in (5.46) by

$$\mathcal{L}^* G = \frac{\mathrm{d}^2 G}{\mathrm{d}x^2} - \frac{\mathrm{d}}{\mathrm{d}x}(p(x)G) + q(x)G,$$

$\mathcal{L}^*$ being the operator adjoint to $\mathcal{L}$.

[67]This is unlike the situation in §4.2, where the characteristic directions were vital.

where D_ϵ is a circle with centre $\boldsymbol{\xi}$ and radius ϵ that has been excised from D. Noting that $\partial/\partial n|_{\partial D_\epsilon} = -\partial/\partial r$ in polar coordinates centred at $\boldsymbol{\xi}$, we find

$$\int_{\partial D_\epsilon}\left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right)\mathrm{d}s = \int_0^{2\pi}\left(u(\boldsymbol{\xi})\left(-\frac{1}{2\pi\epsilon}\right) - \frac{\partial u}{\partial n}(\bar{\boldsymbol{\xi}})\left(\frac{\log\epsilon}{2\pi}\right) + O(1)\right)\epsilon\,\mathrm{d}\theta \tag{5.51}$$

as $\epsilon \to 0$, for some $\bar{\boldsymbol{\xi}}$ such that $\left|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}\right| < \epsilon$. Substituting into (5.50) and using (5.43), (5.48) and the data on ∂D, and taking the limit $\epsilon \to 0$, we end up with our desired formula

$$u(\boldsymbol{\xi}) = \iint_D G(\mathbf{x},\boldsymbol{\xi})f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\partial D} g(\mathbf{x})\frac{\partial G}{\partial n}(\mathbf{x},\boldsymbol{\xi})\,\mathrm{d}s. \tag{5.52}$$

There are two immediate remarks to be made about this result.

- It is easy to see that an equation similar to (5.52) applies to the Robin problem, as long as G satisfies the homogeneous form of the Robin boundary condition. Also it is easy to modify the argument to cater for other second-order elliptic operators, except that when the operator is not self-adjoint, say when

$$\mathcal{L} = \nabla^2 + a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y},$$

  we have to work with the adjoint operator defined by

$$\mathcal{L}^*G = \nabla^2 G - \frac{\partial}{\partial x}(aG) - \frac{\partial}{\partial y}(bG).$$

- Although we have defined Green's functions in a framework analogous to that for Riemann functions, their properties are very different: G depends on the domain D, whereas R is independent of the initial curve. Moreover, G can only vanish at isolated points, whereas R is zero for all $x > \xi$ and all $y > \eta$. The heart of the analogy is revealed by the 'short-cut' approach, which we now describe.

5.5.2 Generalised function formulation

Precisely in line with §4.2.2, we could, instead of writing (5.49), define G to be such that

$$\nabla^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} = \delta(\mathbf{x} - \boldsymbol{\xi}) \quad \text{for } \boldsymbol{\xi} \in D, \tag{5.53}$$

where we recall $\delta(\mathbf{x} - \boldsymbol{\xi}) = \delta(x - \xi)\delta(y - \eta)$ is zero except at $\mathbf{x} = \boldsymbol{\xi}$ and is such that $\iint_D \delta(\mathbf{x} - \boldsymbol{\xi})\,\mathrm{d}\mathbf{x} = 1$. Then, assuming that $G = 0$ on ∂D, we can derive (5.52) in one line by assuming Green's theorem holds for $\iint_D \left(u\nabla^2 G - G\nabla^2 u\right)\mathrm{d}\mathbf{x}$. As usual, we simply multiply (5.53) by u, Poisson's equation $\nabla^2 u = f$ by G, integrate over D and subtract to get

$$\iint_D (u\nabla^2 G - G\nabla^2 u)\,\mathrm{d}\mathbf{x} = \int_{\partial D}\left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right)\mathrm{d}s = \int_{\partial D} g(\mathbf{x})\frac{\partial G}{\partial n}\,\mathrm{d}s$$
$$= \iint_D (u(\mathbf{x})\delta(\mathbf{x}-\boldsymbol{\xi}) - G(\mathbf{x},\boldsymbol{\xi})f(\mathbf{x}))\,\mathrm{d}\mathbf{x}$$
$$= u(\boldsymbol{\xi}) - \iint_D G(\mathbf{x},\boldsymbol{\xi})f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

As discussed in §4.2.2, the justification for the use of Green's theorem is the nub of the matter and further reassurance concerning its validity is given in Exercise 5.9.

We can now set out our general theoretical framework for an elliptic operator $\mathcal{L}$ with adjoint $\mathcal{L}^*$. Everything relies on the fact that, if, as functions of $\mathbf{x}$,

$$\mathcal{L}^*G(\mathbf{x},\boldsymbol{\xi}) = \delta(\mathbf{x}-\boldsymbol{\xi}) \quad \text{and} \quad \mathcal{L}u(\mathbf{x}) = f(\mathbf{x}),$$

then, with homogeneous boundary conditions, the identity

$$\iint_D (G(\mathbf{x},\boldsymbol{\xi})\mathcal{L}u(\mathbf{x}) - u(\mathbf{x})\mathcal{L}^*G(\mathbf{x},\boldsymbol{\xi}))\,\mathrm{d}\mathbf{x} = 0$$

implies that

$$u(\boldsymbol{\xi}) = \iint_D G(\mathbf{x},\boldsymbol{\xi})f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

We can now make some more remarks.

- Equation (5.53) emphasises the ancestry of all Green's functions that we outlined in §4.4. They are such that a linear operator annihilates them at all points except one. Moreover, if the operator $\mathcal{L}$ is self-adjoint with appropriate boundary conditions, then G is *symmetric*, i.e.

  $$G(\mathbf{x},\boldsymbol{\xi}) = G(\boldsymbol{\xi},\mathbf{x}).$$

  To prove this simply set

  $$\mathcal{L}^*G(\mathbf{x},\boldsymbol{\eta}) = \delta(\mathbf{x}-\boldsymbol{\eta})$$

  and use the fact that

  $$\iint_D (G(\mathbf{x},\boldsymbol{\eta})\mathcal{L}G(\mathbf{x},\boldsymbol{\xi}) - G(\mathbf{x},\boldsymbol{\xi})\mathcal{L}^*G(\mathbf{x},\boldsymbol{\eta}))\,\mathrm{d}\mathbf{x} = 0.$$

  This fact is important when reconciling the views of people who boldly assert that the solution of Poisson's equation with zero Dirichlet boundary data is

  $$u(\mathbf{x}) = \iint_D \widetilde{G}(\mathbf{x},\boldsymbol{\xi})f(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}, \tag{5.54}$$

  just as long as $\nabla^2\widetilde{G}(\mathbf{x},\boldsymbol{\xi}) = \delta(\mathbf{x}-\boldsymbol{\xi})$, by simply differentiating under the integral sign. Of course, their $\widetilde{G}$ is simply the 'transpose' of G defined by (5.53) and, for any self-adjoint operator, there is no confusion. Indeed, this argument shows that when $\mathcal{L} \neq \mathcal{L}^*$ the Green's function satisfies $\mathcal{L}G = \delta(\mathbf{x}-\boldsymbol{\xi})$ as a function of $\boldsymbol{\xi}$, as well as $\mathcal{L}^*G = \delta(\mathbf{x}-\boldsymbol{\xi})$ as a function of $\mathbf{x}$.

- When we extend (5.53) to $\mathbb{R}^m$, we simply find that the singular part of (5.49) is replaced by
$$-\frac{\omega_m}{r^{m-2}} \quad \text{for } m \geqslant 3, \tag{5.55}$$
where ω_m is the surface area of the unit sphere in $\mathbb{R}^m$, namely[68]
$$\omega_m = \begin{cases} m\pi^{m/2}/(m/2)!\,, & m \text{ even}, \\ \pi^{(m-1)/2}2^m\,((m-1)/2)!/(m-1)!\,, & m \text{ odd}. \end{cases}$$
If D is infinite and $u \to 0$ at infinity, G is precisely this function and it models a point 'charge' or point 'mass'; this is just our comment before (5.1). When $m = 2$, a physical interpretation of (5.49) is as the electrostatic field due to a line charge as mentioned in §5.1.2. We also remark that the singularity in the Green's function of an arbitrary elliptic operator takes the form above after the equation has been put into canonical form.
- The whole question of the existence of G is as vexed as that of the existence of solutions for general elliptic equations, discussed briefly at the end of the last section.[69] Of course, if G does exist, formulæ such as (5.52) provide a ready tool to demonstrate the continuous dependence of the solution on the data. However, there is one commonly occurring situation where G definitely does not exist and the procedures above must be modified. Suppose we try to solve the Neumann problem for Poisson's equation $\nabla^2 u = f$ as in (5.35), with $\partial u/\partial n = g(\mathbf{x})$ on ∂D, and we know that (5.39) is satisfied. We know at once that G only exists if the term that forces G to be non-zero, namely the right-hand side $\delta(\mathbf{x} - \boldsymbol{\xi})$ of (5.53), is orthogonal to the constant eigenfunction. But $\delta(\mathbf{x} - \boldsymbol{\xi})$ has a unit integral so this can never happen. Fortunately, there are at least two ways out of this difficulty. One is to give G an extra degree of freedom by choosing
$$\nabla^2 G_M = \delta(\mathbf{x} - \boldsymbol{\xi}) + c\delta(\mathbf{x} - \boldsymbol{\eta}),$$
pick $c = -1$ to satisfy orthogonality, and arrive at
$$u(\boldsymbol{\xi}) - u(\boldsymbol{\eta}) = \iint_D G_M(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\eta}) f(\mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\partial D} g(\mathbf{x}) G_M(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\eta})\,\mathrm{d}s; \tag{5.56}$$
the right-hand side is easily seen to be the sum of a function of $\boldsymbol{\xi}$ and a function of $\boldsymbol{\eta}$. Another way is to choose

[68]To prove this quickly, note that
$$\pi^{m/2} = \int_{\mathbb{R}^m} \mathrm{e}^{-r^2}\,\mathrm{d}\mathbf{x} = \omega_m \int_0^\infty r^{m-1}\mathrm{e}^{-r^2}\,\mathrm{d}r, \quad \text{where } r = |\mathbf{x}|.$$

[69]We remark that, if we have proved the existence of the solution of a boundary value problem for an elliptic equation, we can use this information to deduce the existence of a Green's function. For example, for the Dirichlet problem for Laplace's equation in two dimensions, $G = (1/2\pi)\log|\mathbf{x} - \boldsymbol{\xi}| + G_R(\mathbf{x}, \boldsymbol{\xi})$, where G_R, the *regular part* of the Green's function, satisfies the boundary value problem $\nabla^2 G_R = 0$ in D and $G_R = -(1/2\pi)\log|\mathbf{x} - \boldsymbol{\xi}|$ on ∂D.

$$\nabla^2 G_M = \delta(\mathbf{x} - \boldsymbol{\xi}) + c \cdot 1, \tag{5.57}$$

using the fact that 1 is an eigensolution. Now pick $-1/c$ to be the area of D to ensure orthogonality and a formula similar to (5.56) emerges.

- As mentioned after (5.52), our definition of G easily generalises to the Robin problem where (5.36) holds with neither β nor α equal to zero. If u satisfies Poisson's equation with this condition, we simply set $\nabla^2 G = \delta(\mathbf{x} - \boldsymbol{\xi})$ with $\alpha G + \beta\, \partial G/\partial n = 0$ on ∂D to ensure that $\int_{\partial D} (G\, \partial u/\partial n - u\, \partial G/\partial n)\, ds$ can be written in terms of G and g alone.

There is one other very important remark to be made about Green's functions from the computational viewpoint. This is that, if we relax the boundary conditions that we have imposed on G, we can always use (5.50) to relate the values of u in D to integrals involving both u and $\partial u/\partial n$ around ∂D. Hence, by taking a suitable limiting process, as in Exercise 5.13, we can derive a linear integral equation for the values of u or $\partial u/\partial n$ on ∂D for any of the boundary value problems involving (5.36). The form of this integral equation depends on the problem and the way in which we manipulate (5.50), but it is the basis of the so-called ***boundary integral method*** [6] for solving Laplace's equation and many other elliptic equations. Discretising the integral equation over the whole of ∂D gives fewer linear algebraic equations to be solved than would be necessary with a conventional finite difference or finite element discretisation, but the other side of the coin is that the matrices resulting from the boundary integral discretisation are 'full', in that they have few zero entries.

It is with sadness that we must tell the reader that explicit formulæ for Green's functions rarely exist (see [34] for a catalogue). However, things seem a little better than they are for Riemann functions; although we can only cope with a small handful of equations, there are quite a few geometries that have enough symmetry for us to be able to find G.

5.6 Explicit representations of Green's functions

5.6.1 Laplace's equation and Poisson's equation

We have seen that the Green's functions that are as well-behaved as possible at infinity in $\mathbb{R}^m$ are $(1/2\pi)\log r$ and $-1/4\pi r$ for $m = 2$ and $m = 3$, respectively. Things rapidly become more complicated when boundaries are introduced, as we now see.

5.6.1.1 Circles and spheres: Dirichlet and Neumann conditions

Suppose D is a circle with centre $\mathbf{x} = \mathbf{0}$ and radius a in $\mathbb{R}^2$. For the geometrically minded, we can proceed by defining the point inverse to $\boldsymbol{\xi}$ by $\boldsymbol{\xi}' = a^2\boldsymbol{\xi}/|\boldsymbol{\xi}|^2$ and set $R = |\mathbf{x} - \boldsymbol{\xi}|$ and $R' = |\mathbf{x} - \boldsymbol{\xi}'|$, as in Fig. 5.4.[70] Then $R/R' = |\boldsymbol{\xi}|/a$ when $r = |\mathbf{x}| = a$, and $\log R$ and $\log R'$ are both solutions of Laplace's equation since changing

[70] When X lies on the circle, triangles XOP and QOX are similar because they have a common angle $\theta - \alpha$ between sides in the same ratio, as $OQ/OX = OX/OP$ by the image condition; hence, $R = R'|\boldsymbol{\xi}|/a$ when $|\mathbf{x}| = a$.

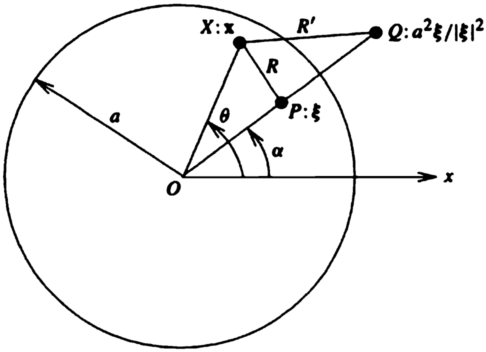


Fig. 5.4 Image points in a circle or sphere.

the origin does not change the equation. Thus, recalling that $G - (1/2\pi)\log R$ is bounded as $R \to 0$, and since $G = 0$ on $r = a$, the formula for G is

$$G = \frac{1}{2\pi}\left(\log R - \log\left(R'\frac{|\boldsymbol{\xi}|}{a}\right)\right). \tag{5.58}$$

This is an example of the *method of images*, $\boldsymbol{\xi}'$ being the image of $\boldsymbol{\xi}$ in $r = a$. With reference to electrostatics, the idea is that the field generated by a line charge inside a perfect conductor in the form of a circular cylinder is the same as that with the conductor replaced by an equal and opposite line charge through the inverse point; this idea of replacing boundary conditions by appropriate distributions of singularities is one to which we will return several times. To calculate $u(\boldsymbol{\xi})$ we need to let α be the polar angle of $\boldsymbol{\xi}$ and calculate

$$\begin{aligned}\left.\frac{\partial G}{\partial n}\right|_{r=a} &= \left.\frac{\partial G}{\partial r}\right|_{r=a}\\ &= \frac{1}{4\pi R^2}\frac{\partial}{\partial r}\left(r^2 + |\boldsymbol{\xi}|^2 - 2r|\boldsymbol{\xi}|\cos(\theta-\alpha)\right)\Big|_{r=a}\\ &\quad - \frac{1}{4\pi R'^2}\frac{\partial}{\partial r}\left(r^2 + \frac{a^4}{|\boldsymbol{\xi}|^2} - 2r\frac{a^2}{|\boldsymbol{\xi}|}\cos(\theta-\alpha)\right)\Bigg|_{r=a}\\ &= \frac{a^2 - |\boldsymbol{\xi}|^2}{2\pi a R^2},\end{aligned}$$

since $R/R' = |\boldsymbol{\xi}|/a$. Hence, when $f = 0$, we derive the famous *Poisson integral*

$$u(\boldsymbol{\xi}) = \frac{a^2 - |\boldsymbol{\xi}|^2}{2\pi}\int_0^{2\pi}\frac{g(\theta)\,\mathrm{d}\theta}{|\boldsymbol{\xi}|^2 + a^2 - 2a|\boldsymbol{\xi}|\cos(\theta-\alpha)}. \tag{5.59}$$

This method also works if we consider the region exterior to $r = a$, with one important proviso. Since the region is both infinite and multiply connected we may need to take especial care to exclude eigensolutions. For the Dirichlet problem, it is sufficient to say that $u = o(\log r)$ as $r \to \infty$ so as to exclude the eigensolutions

$$\log \frac{r}{a} \quad \text{and} \quad \left(r^n - \frac{a^{2n}}{r^n}\right)(A\cos n\theta + B\sin n\theta),$$

where n is an integer greater than zero, and A and B are constants.

No formula such as (5.59) is available for the Robin condition (5.13), but the Green's function for the Neumann problem can still be represented by images. We will leave the details to Exercise 5.17, where it is shown that when $|\boldsymbol{\xi}| > a$ the image system comprises *two* line charges inside the circle when we exclude eigensolutions for which $u = \text{constant} \cdot \theta + o(1)$ as $r \to \infty$. On the other hand, if $|\boldsymbol{\xi}| < a$, we know there is no usual Green's function and we have to resort to (5.56); the formula quoted in Exercise 5.18 is an example of this.

Since Fig. 5.4 could equally well be drawn in $\mathbb{R}^3$, the ideas of inversion can often be used for spheres as well as for circles, with $G = -1/4\pi r + O(1)$ near the singularity in spherical polar coordinates.[71]

For the next class of problems, Green's functions can be found by 'turning a handle' (in fact, by taking Fourier transforms).

5.6.1.2 Half-spaces

Suppose we have to solve Laplace's equation in $y > 0$ in $\mathbb{R}^2$, with Dirichlet data $u = g_D(x)$ on $y = 0$ and suitable behaviour at infinity to ensure uniqueness. Then we can either use images directly, or take the limit of (5.58), to find that

$$G = \frac{1}{2\pi}\log\frac{|\mathbf{x}-\boldsymbol{\xi}|}{|\mathbf{x}-\boldsymbol{\xi}'|},$$

so that the solution, which we write as[72] $u_D(x, y)$, is

$$u_D(x,y) = \frac{y}{\pi}\int_{-\infty}^{\infty}\frac{g_D(\xi)\,\mathrm{d}\xi}{(x-\xi)^2+y^2}. \tag{5.60}$$

Note that $u_D(x, y)$, as defined by (5.60), also satisfies Laplace's equation for $y < 0$; its normal derivative is continuous across $y = 0$, but u_D itself has a jump of $2g_D$ there. It also shows that the behaviour at infinity that we need is $u = O(1/r)$ as $r^2 = x^2 + y^2 \to \infty$.

We could have derived this result more methodically by introducing the Fourier transform of u_D,

$$\hat{u}_D(k,y) = \int_{-\infty}^{\infty} u_D(x,y)\mathrm{e}^{ikx}\,\mathrm{d}x,$$

[71] We will not pursue this here, but there is another mysterious-looking connection between inversion and Laplace operators in $\mathbb{R}^m$. This is that, if $u(r, \theta_1, \ldots, \theta_{m-1})$ is a harmonic function in $\mathbb{R}^m$, with r being distance from the origin and $\theta_1, \ldots, \theta_{m-1}$ being polar angles, then so is $r^{2-m}u(a^2/r, \theta_1, \ldots, \theta_{m-1})$. This result basically comes from separating the variables in polar coordinates and noting that powers of r always appear as pairs of the form r^λ and $r^{-\lambda-m+2}$.

[72] In this section, for ease of presentation, we use the Green's function to find $u(\mathbf{x})$ rather than $u(\boldsymbol{\xi})$ as hitherto. There should be no confusion as G is symmetric.

using the postulated decay at infinity, to obtain

$$\frac{d^2\hat{u}_D}{dy^2} - k^2\hat{u}_D = 0, \tag{5.61}$$

with

$$\hat{u}_D(k,0) = \hat{g}_D(k) = \int_{-\infty}^{\infty} g_D(x)e^{ikx}\,dx.$$

Hence, assuming we restrict ourselves to real k,

$$\hat{u}_D(k,y) = \hat{g}_D(k)e^{-|k|y} = \hat{g}_D(k)\hat{H}(k,y), \tag{5.62}$$

say. Now the Fourier inversion of $\hat{H}$ can be performed on the real axis to give

$$H(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx-|k|y}\,dk = \frac{1}{\pi}\int_0^{\infty} e^{-ky}\cos kx\,dk = \frac{y}{\pi(x^2+y^2)}.$$

Hence (5.60) follows by the convolution theorem; clearly

$$H(x-\xi,y) = -\left.\frac{\partial G}{\partial \eta}\right|_{\eta=0}.$$

This method has the advantage of working for Neumann and Robin as well as Dirichlet data, and in principle it can be used for any constant-coefficient elliptic equation in a half-space.

If $u_N(x,y)$ is the solution of the Neumann problem in $y > 0$, in which $\partial u_N/\partial y\,(x,0) = g_N(x)$ is given, it is easy to show that a Green's function is

$$G = \frac{1}{2\pi}\left(\log|\mathbf{x}-\boldsymbol{\xi}| + \log|\mathbf{x}-\boldsymbol{\xi}'|\right),$$

so that

$$u_N(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_N(\xi)\log\left((x-\xi)^2+y^2\right)d\xi, \tag{5.63}$$

and an arbitrary constant can be added. Note that $u_N(x,y)$ can only be bounded at infinity if $\int_{-\infty}^{\infty} g_N(\xi)\,d\xi = 0$, which is the analogue of the usual solvability condition in a finite domain. Note also that $u_N(x,y)$ satisfies Laplace's equation for $y < 0$ and is continuous across $y = 0$, but that its normal derivative has a jump of $2g_N(x)$. Moreover, this solution, which is a superposition of Green's functions that are singular on the x axis, has a simple interpretation as a 'source distribution' in fluid mechanics, gravitation or electromagnetism.

Still in connection with the half-plane geometry, there is one very useful piece of jargon that can be introduced. It is easy to see that the Fourier transform of u_N is

$$\hat{u}_N(k,y) = -\frac{\hat{g}_N(k)}{|k|}e^{-|k|y},$$

and hence the Dirichlet and Neumann data for u are related by

$$\hat{g}_N(k) = -|k|\,\hat{g}_D(k). \tag{5.64}$$

This is an example of a so-called *Dirichlet-to-Neumann map* and it sheds further light on the ill-posedness of the Cauchy problem for Laplace's equation in $y > 0$.

By the Cauchy–Kowalevski theorem, there is a unique local solution to this Cauchy problem if $\hat{g}_D$ and $\hat{g}_N$ are the transforms of arbitrary analytic functions, but the solution only exists globally in $y > 0$ and has the required properties at infinity if $\hat{g}_D$ and $\hat{g}_N$ are related by (5.64). To interpret this result, one procedure is to note that g_D is the convolution of g_N and the Fourier inverse of $-1/|k|$. Unfortunately, the latter is difficult to invert although, if we are daring, we can show that $\mathrm{i}k/|k|$ is the Fourier transform of $1/\pi x$ (but we must use the arguments of §4.2.2 and then wait for Exercise 5.21 to see this). This suggests that $\mathrm{d}g_D/\mathrm{d}x$, whose transform is $-\mathrm{i}k\hat{g}_D$, is the convolution of g_N and $1/\pi x$, which tempts us to write

$$\frac{\mathrm{d}g_D}{\mathrm{d}x} = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g_N(\xi)\,\mathrm{d}\xi}{\xi - x}, \tag{5.65}$$

providing we can make sense of the integral. A precisely similar argument for (5.64), rewritten as

$$\hat{g}_N(k) = \mathrm{i}k\left(\frac{\mathrm{i}|k|}{k}\right)\hat{g}_D(k),$$

yields

$$g_N(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{d}g_D}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x}, \tag{5.66}$$

but again we need to interpret the integral. Fortunately, we can revert to (5.60) and integrate by parts to write it as

$$u_D(x, y) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \tan^{-1}\left(\frac{\xi - x}{y}\right)\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\,\mathrm{d}\xi$$

for $y > 0$, where we must define the inverse tangent as lying between $-\pi/2$ and $\pi/2$. As $y \downarrow 0$,

$$\tan^{-1}\left(\frac{\xi - x}{y}\right) = \pm\frac{\pi}{2} - \frac{y}{\xi - x} + o(y)$$

for $\xi - x > 0$ and $\xi - x < 0$, respectively. Hence, in order to find the limiting behaviour of $u_D(x, y)$ as $y \downarrow 0$, we need to split the integral into $\lim_{\epsilon\to 0}\left(\int_{-\infty}^{x-\epsilon} + \int_{x+\epsilon}^{\infty}\right)$ so that, when y is small,

$$\begin{aligned} u_D(x, y) &= \lim_{\epsilon\to 0}\left(\frac{1}{\pi}\int_{-\infty}^{x-\epsilon}\left(\frac{\pi}{2} + \frac{y}{\xi - x} + \cdots\right)\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\,\mathrm{d}\xi\right. \\ &\qquad\left. + \frac{1}{\pi}\int_{x+\epsilon}^{\infty}\left(-\frac{\pi}{2} + \frac{y}{\xi - x} + \cdots\right)\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\,\mathrm{d}\xi\right) \\ &= g_D(x) + \frac{y}{\pi}\mathrm{PV}\int_{-\infty}^{\infty}\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x} + \cdots, \end{aligned}$$

where

$$\mathrm{PV}\int_{-\infty}^{\infty}\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x} = \lim_{\epsilon\to 0}\left(\int_{-\infty}^{x-\epsilon} + \int_{x+\epsilon}^{\infty}\right)\frac{\mathrm{d}g_D}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x} \tag{5.67}$$

is called a ***principal value integral***. We can now differentiate with respect to y and set $y = 0$ to find

$$g_N(x) = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty} \frac{\mathrm{d}g_D}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x}, \tag{5.68}$$

which tells us how to interpret the integrals above. The right-hand side of (5.68) is called the ***Hilbert transform*** of $\mathrm{d}g_D/\mathrm{d}\xi$.

5.6.1.3 Strips and rectangles

It is a simple exercise to repeat the transform calculations above in the strip $0 < y < 1$, $-\infty < x < \infty$, say, with $u(x,0) = 0$, $u(x,1) = g(x)$ and $u \to 0$ as $|x| \to \infty$. We obtain

$$u(x,y) = \int_{-\infty}^{\infty} g(\xi)H(\xi - x, y)\,\mathrm{d}\xi, \tag{5.69}$$

where now

$$H(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ikx}\frac{\sinh ky}{\sinh k}\,\mathrm{d}k.$$

The function H is again the y derivative of the Green's function evaluated at $\eta = 1$ and, by writing it as an infinite series of contributions from poles at $k = in\pi$, $n = 0, \pm 1, \ldots$, we can interpret the Green's function in terms of an infinitely repeated series of images in $y = 0$ and $y = 1$.

It is sometimes possible to solve problems in semi-infinite strips by using a different kind of image. Suppose that u is harmonic in the half-strip $0 < y < \infty$, $0 < x < 1$, and that

$$u(x,0) = g(x) \quad \text{for } 0 < x < 1, \qquad u(0,y) = u(1,y) = 0 \quad \text{for } y > 0,$$

together with boundedness conditions at $(0,0)$ and $(1,0)$. We simply extend the problem periodically to $-\infty < x < \infty$ with u odd in x; then (5.60) gives

$$u(x,y) = \frac{y}{\pi}\sum_{n=-\infty}^{\infty}\int_0^1 g(\xi)\left(\frac{1}{(x-2n-\xi)^2+y^2} - \frac{1}{(x-(2n+1)-\xi)^2+y^2}\right)\mathrm{d}\xi, \tag{5.70}$$

which is much quicker than using Green's functions. We could also have proceeded by separating the variables, to obtain a series which converges well for large y, in contrast to (5.70), which works well for small y.

To those who find the prospect of similarly constructing series of images in a rectangle daunting, the following technique for finding Green's functions will come as a relief; in fact it works in any bounded domain for which we know the eigenfunctions of the Laplacian. Suppose we want to solve the Dirichlet problem for Laplace's equation in $-a < x < a$, $-b < y < b$, with, of course, appropriate boundedness conditions at each vertex. As usual,

$$\nabla^2 G = \delta(x - \xi)\delta(y - \eta), \tag{5.71}$$

with $G = 0$ on the boundary.[73] What we now do is define the *finite Fourier transform* of G, which is a generalisation of the Fourier series we discussed in §4.4. There we expanded arbitrary functions in terms of complete sets of eigenfunctions of the appropriate eigenvalue problems; here for an arbitrary function $F(x,y)$ we simply define

$$\hat{F}_{mn} = \int_{-b}^{b}\int_{-a}^{a} F(x,y)\sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}\,dx\,dy, \tag{5.72}$$

where the kernel is just an eigenfunction of the Laplacian with zero Dirichlet conditions, and corresponding eigenvalue $-(m^2\pi^2/a^2+n^2\pi^2/b^2)$. Multiplying both sides of (5.71) by this kernel and integrating gives

$$\hat{G}_{mn} = -\sin\frac{m\pi\xi}{a}\sin\frac{n\pi\eta}{b}\Big/\left(\frac{m^2\pi^2}{a^2}+\frac{n^2\pi^2}{b^2}\right).$$

The inversion formula for (5.72) in this case is exactly that for Fourier series, namely

$$G = \frac{1}{ab}\sum_{m,n}\hat{G}_{mn}\sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}. \tag{5.73}$$

Of course, (5.72) can be generalised and used to obtain G in a general domain provided that we know the eigenfunctions of the Laplacian in that domain. We return to this question in §5.7.2.

5.6.2 Helmholtz' equation

Because of its importance in diffraction theory, we will accord Helmholtz' equation a special section. We will restrict ourselves to two dimensions for simplicity, and we see that the Green's function for the whole plane satisfies

$$(\nabla^2 + k^2)\,G = \delta(\mathbf{x}-\boldsymbol{\xi}). \tag{5.74}$$

By writing this in polar coordinates, we see that, instead of G being a combination of $\log r$ and a constant as it was for Laplace's equation, it is now a combination of the Bessel functions $J_0(kr)$ and $Y_0(kr)$. Clearly, we need G to have a logarithmic singularity with strength $1/2\pi$ as $r = |\mathbf{x}-\boldsymbol{\xi}| \to 0$, which, from known properties of Y_0, demands that $G = (1/4)Y_0(kr) + \text{constant}\cdot J_0(kr)$. However, we need to specify the behaviour as $r \to \infty$ more carefully before we can proceed. In particular, we need to prescribe how much of the solution represents an 'outgoing' wave as $r \to \infty$. To do this we recall from Chapter 4 that waves in which $u \sim f(r + a_0 t)$ as $r \to \infty$ are incoming and those in which $u \sim f(r - a_0 t)$ are outgoing.[74] Writing

[73] In fact, the problem for G is a good model for certain industrial processes involving current flow in a conducting plate to which an electrode is attached at $x = \xi$, $y = \eta$; G is simply the electric potential.

[74] The functions $f(r \pm a_0 t)$ are not, of course, exact solutions of the wave equation except in one space dimension, but as $r \to \infty$ we expect arbitrary waves to become more and more one-dimensional.

$$f \sim e^{ik(r \pm a_0 t)},$$

and remembering the factor $e^{-i\omega t}$ in our derivation of (5.74), we obtain the motivation for a *radiation condition* for Helmholtz' equation in the form

$$\frac{\partial G}{\partial r} - ikG = o(G)$$

as $r \to \infty$, which is just a concise way of writing that G tends to e^{+ikr} multiplied by some function which may decay algebraically in r, rather than e^{-ikr} multiplied by such a function. In fact, it can be shown by using asymptotic methods (see Exercise 8.11) that this radiation condition can be written more precisely as the *Sommerfeld radiation condition*

$$\lim_{r \to \infty} r^{1/2} \left(\frac{\partial G}{\partial r} - ikG \right) = 0. \tag{5.75}$$

A quick glance at the asymptotic expansions of J_0 and Y_0 [8] shows that the only combination that satisfies (5.75) is a multiple of the *Hankel function*

$$H_0^{(1)}(kr) = J_0(kr) + iY_0(kr). \tag{5.76}$$

Thus

$$G = -\frac{i}{4} H_0^{(1)} \left(k \left| \mathbf{x} - \boldsymbol{\xi} \right| \right), \tag{5.77}$$

which, as expected from §5.1.5, is complex; it is still symmetric in $\mathbf{x}$ and $\boldsymbol{\xi}$ because $\nabla^2 + k^2$ is self-adjoint.

It would be sadistic to inflict on the reader a list of Green's functions for Helmholtz' equation in the presence of boundaries, even though there are precious few of them; one example is that of a rectangle with zero Dirichlet data, where we can use a simple modification of (5.73), as long as $-k^2$ is not an eigenvalue of the Laplacian (see Exercise 5.16). However, we can point out that, although the presence of finite boundaries in an otherwise infinite domain requires that G should, as usual, satisfy appropriate homogeneous boundary conditions on them, the radiation condition (5.75) remains unaltered.[75] We simply cite what can happen if the geometry is simple enough. Suppose, for example, that

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + k^2 u = 0$$

[75] It can be shown that the presence of finite boundaries merely introduces a 'directivity' $A(\theta_1, \ldots, \theta_{m-1})$ into the far-field so that

$$G \sim \frac{A(\theta_1, \ldots, \theta_{m-1}) e^{ikr}}{r^{(m-1)/2}},$$

where m is the dimensionality and $\theta_1, \ldots, \theta_{m-1}$ are polar angles (see Chapter 8); this makes an interesting contrast with the far-field of the Green's function for Laplace's equation, which is isotropic in general.

in polar coordinates, with $u = f(\theta)$ on $r = 1$. Then the solution in $r > 1$ that satisfies (5.75) is

$$u(r,\theta) = \Re \sum_{n=0}^{\infty} (a_n \cos n\theta + b_n \sin n\theta)\, H_n^{(1)}(kr),$$

and the solution in $r < 1$ is

$$u(r,\theta) = \sum_{n=0}^{\infty} (c_n \cos n\theta + d_n \sin n\theta)\, J_n(kr).$$

Here a_n, b_n, c_n and d_n are easily related to the Fourier coefficients of $f(\theta)$. Also $J_n(x)$, the Bessel function of the first kind of order n, satisfies

$$\frac{\mathrm{d}^2 J_n}{\mathrm{d}x^2} + \frac{1}{x}\frac{\mathrm{d}J_n}{\mathrm{d}x} + \left(1 - \frac{n^2}{x^2}\right) J_n = 0, \qquad J_n(x) = \frac{x^n}{2^n n!}\left(1 + O(x^2)\right) \qquad \text{as } x \to 0, \tag{5.78}$$

while $Y_n(x)$, the Bessel function of the second kind, is equal to

$$\frac{J_n(x)\cos n\pi - J_{-n}(x)}{\sin n\pi}$$

for non-integral n, and to the limit of this expression for integral n; it is singular at the origin (for more details see [8]); the combination $H_n^{(1)} = J_n + \mathrm{i}Y_n$ is known as a Hankel function of order n. For future reference, we note that $J_n(x)$ is proportional to $x^n \int_0^\pi \cos(x\cos\theta)\sin^{2n}\theta\,\mathrm{d}\theta$; this result is true even for n complex.

Lastly, we note that the situation becomes more complicated when the boundaries extend to infinity because the condition for outgoing waves may have to be modified.

5.6.3 The modified Helmholtz equation

The modified Helmholtz equation,

$$(\nabla^2 - k^2)u = 0,$$

has quite different properties from the ordinary Helmholtz equation, because in an unbounded domain its solutions either grow or decay exponentially at infinity, rather than algebraically. Its Green's function in $\mathbb{R}^2$ is the solution of

$$\frac{\mathrm{d}^2 G}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}G}{\mathrm{d}r} - k^2 G = 0$$

that decays at infinity and has the correct logarithmic behaviour at the origin, namely $G = -(1/2\pi)K_0(kr)$, where $K_0(kr)$ is a modified Bessel function of the third kind, equal to $(\mathrm{i}\pi/2)H_0^{(1)}(\mathrm{i}kr)$. It is left to the reader to show that the Green's function in $\mathbb{R}^3$ is $-\mathrm{e}^{-kr}/4\pi r$; note that neither of these Green's functions is oscillatory.

*5.7 Green's functions, eigenfunction expansions and transforms

Results such as the series (5.73) suggest that we look more closely at the relationships between Green's functions, eigenfunctions and transforms. They motivate thinking of the Green's function $G(\mathbf{x}, \boldsymbol{\xi})$ for a self-adjoint operator $\mathcal{L}$ as a superposition of eigenfunctions $\phi_\lambda(\mathbf{x})$ that satisfy $\mathcal{L}\phi_\lambda = \lambda\phi_\lambda$, and then regarding our basic formula (5.52) as a transform formula in the sense of §4.4. Put another way, the solution of $\mathcal{L}u = f$ emerges as a superposition of ϕ_λ whose coefficients are easy to find because

$$\int \phi_\lambda f \,\mathrm{d}\mathbf{x} = \int \phi_\lambda \mathcal{L}u \,\mathrm{d}\mathbf{x} = \int u\mathcal{L}\phi_\lambda \,\mathrm{d}\mathbf{x} = \lambda \int u\phi_\lambda \,\mathrm{d}\mathbf{x}.$$

Each elliptic problem can thus be solved by its customised transform, taken with respect to its eigenfunctions ϕ_λ, and this procedure can be modified to cater for non-self-adjoint operators. However, before we can describe the method we need to review a few important results about eigenfunctions and eigenvalues of elliptic operators.

5.7.1 Eigenvalues and eigenfunctions

Most of what we need to know about this large subject is a generalisation of the Sturm–Liouville theory for two-point boundary value problems for self-adjoint second-order ordinary differential equations [44]. In that theory, a principal result is that the eigenvalues are real and the eigenfunctions are complete and can be orthonormalised, those corresponding to different eigenvalues being orthogonal automatically. The orthogonality results are proved in the same way as for matrices but the completeness is more complicated. One relatively painless way to generalise this procedure to a Dirichlet problem in more dimensions is to note that, if we define the bilinear form $a(u, v)$ of two suitable functions u and v by

$$a(u, v) = \int_D u\mathcal{L}v \,\mathrm{d}\mathbf{x} = \int_D v\mathcal{L}u \,\mathrm{d}\mathbf{x}$$

for a self-adjoint elliptic operator $\mathcal{L}$ defined in a domain D, then the *Rayleigh quotient*

$$\int_D a(u, u) \,\mathrm{d}\mathbf{x} \Big/ \int_D |u|^2 \,\mathrm{d}\mathbf{x}, \qquad \text{where} \qquad u = 0 \quad \text{on } \partial D,$$

is minimised[76] by the eigenfunction ϕ_0 corresponding to the lowest (smallest in modulus) or *principal* eigenvalue of

$$\mathcal{L}u = \lambda u, \qquad u = 0 \quad \text{on } \partial D.$$

Moreover, the $(n+1)$th eigenfunction ϕ_n can be found by carrying out the same minimisation over test functions u that are orthogonal to their n predecessors.

[76]This is an easy exercise in the calculus of variations (see Exercise 5.15).

Completeness can then be proved by using this minimisation property to show that the remainder after n terms in a Fourier series type of expansion is bounded by the inverse of the nth eigenvalue, in the mean-square sense (see [48]). This procedure also works for Neumann and Robin problems, assuming the correct sign in the latter boundary condition.

Another result of great practical importance is that the lowest eigenfunction of the Dirichlet eigenvalue problem has one sign. For second-order ordinary differential equations this is proved relatively easily by considering the way in which the zeros of a function $u(x)$ such that $\mathcal{L}u = \lambda u$, $u(0) = 0$, 'bunch up' near $x = 0$ as λ increases. However, this idea does not work for elliptic equations and, instead, we set ourselves the task of showing that the zeros (or *nodal lines*) of the $(n+1)$th eigenfunction ϕ_n divide D into no more than $n+1$ subdomains. As explained in detail in [12, 21], we assume the contrary and suppose there are $m > n+1$ such subdomains. Then, picking any $n+1$ of these and denoting by ϕ_{ni} the restriction of the eigenfunction to the ith of these $n+1$ domains (i.e. ϕ_{ni} is ϕ_n in the ith subdomain and zero elsewhere), we can construct a test function

$$u = \sum_{i=0}^{n} a_i \phi_{ni}$$

that is normalised and also orthogonal to the preceding eigenfunctions $\phi_0, \ldots, \phi_{n-1}$. For this test function, the Rayleigh quotient becomes

$$\sum_{i=0}^{n} a_i^2 \int_D \phi_{ni} \mathcal{L} \phi_{ni} \,\mathrm{dx} \Big/ \sum_{i=0}^{n} a_i^2 \phi_{ni}^2 \,\mathrm{dx} = \lambda_n,$$

since $\mathcal{L}\phi_{ni} = \lambda_n \phi_{ni}$. This does not prove rigorously, but it does strongly suggest, that u is the $(n+1)$th eigenfunction; but we know it cannot be because u has singularities on n nodal lines. We hope this contradiction will give the reader enough confidence henceforth to assert that the principal eigenfunction has one sign.

5.7.2 Green's functions and transforms

With this background, let us now embark on our discussion of Green's functions and transforms. The series (5.73) for the Green's function of a rectangle suggests the following result. Suppose the spectrum λ_i of a self-adjoint elliptic operator $\mathcal{L}$ is discrete,[77] that zero is not an eigenvalue,[78] and that the real orthonormalised eigenfunctions are $\phi_n(\mathbf{x})$; then

$$G(\mathbf{x}, \boldsymbol{\xi}) = \sum_n \frac{\phi_n(\mathbf{x})\phi_n(\boldsymbol{\xi})}{\lambda_n}. \tag{5.79}$$

This result can be derived by writing G as $\sum_n c_n(\boldsymbol{\xi})\phi_n(\mathbf{x})$ and observing that $\mathcal{L}G = \sum_n c_n(\boldsymbol{\xi})\lambda_n\phi_n(\mathbf{x})$ can only equal $\delta(\mathbf{x} - \boldsymbol{\xi})$ if $c_m(\boldsymbol{\xi})\lambda_m = \phi_m(\boldsymbol{\xi})$; we simply

[77] This usually means D is bounded, ∂D is not too irregular and the operator has no unbounded coefficients when written in canonical form.

[78] When zero is an eigenvalue, we must employ the devices used to introduce generalised Green's functions in §5.5.

multiply each side by $\phi_m(\mathbf{x})$ and integrate. Moreover, either by using (5.79) with $\lambda = 0$ assumed not to be an eigenvalue, or by directly taking a finite Fourier transform, the solution of $\mathcal{L}u = f$ is

$$u(\xi) = \sum_n \left(\int_D f(\mathbf{x})\phi_n(\mathbf{x})\,\mathrm{d}\mathbf{x} \right) \frac{\phi_n(\xi)}{\lambda_n}. \tag{5.80}$$

This formula is fundamental to our discussion in the next section, but before we put it into practice we need to make some remarks about normalisation and how to cater for complex eigenfunctions.

We assume we have a self-adjoint operator $\mathcal{L}$ and boundary or boundedness conditions such that the normalised eigenfunctions[79] ϕ_λ satisfying $\mathcal{L}\phi_\lambda = \lambda\phi_\lambda$ are complete (when it is convenient, we will use ϕ_λ to denote what we called $\phi(x, k)$ in (4.26), where λ and k are related in a known way). This means that an arbitrary function $f(\mathbf{x})$ can be expanded, either as a series $\sum_\lambda \hat{f}_\lambda\phi_\lambda(\mathbf{x})$, or an integral $\int \hat{f}_\lambda\phi_\lambda(\mathbf{x})\,\mathrm{d}\lambda$, or a combination of both, depending on whether the spectrum is discrete, continuous, or both, respectively. In any case, if the eigenfunctions are orthonormal, we expect that

$$\hat{f}_\lambda = \int_D f\phi_\lambda\,\mathrm{d}\mathbf{x}; \tag{5.81}$$

this is just the familiar Fourier series result in the discrete case and usually the continuous spectrum result can be derived by taking a limit such as $L \to \infty$ when the Fourier series is on the interval $(-L, L)$. However, we must be careful about what we mean by normalisation when we take this limit and, to illustrate this important point, let us return to our Fourier transform discussion in §4.4. Suppose we take $\mathcal{L} = \mathrm{d}^2/\mathrm{d}x^2$, $0 < x < \infty$, and we seek eigenfunctions that vanish at $x = 0$ and are bounded as $x \to \infty$. The only possibility is that λ is a negative number, in which case ϕ_λ is proportional to $\sin kx$, $k = \sqrt{-\lambda}$; but what is the normalisation? To answer this question, we specialise the Fourier transform and its inverse to odd functions to show that

$$u(x) = \frac{2}{\pi}\int_0^\infty \left(\int_0^\infty u(\xi)\sin k\xi\,\mathrm{d}\xi \right) \sin kx\,\mathrm{d}k.$$

Now, this double integral is, formally,

$$\frac{2}{\pi}\int_0^\infty \int_0^\infty u(\xi)\sin k\xi \sin kx\,\mathrm{d}k\,\mathrm{d}\xi$$

and so we can make the interpretation

$$\int_0^\infty \sin k\xi \sin kx\,\mathrm{d}k = \frac{\pi}{2}\delta(x - \xi). \tag{5.82}$$

[79]We denote the spectrum by the single scalar λ even though λ is parametrised by m numbers when there are m independent variables $\mathbf{x}$.

Thus the normalised eigenfunctions are $\phi(x,k) = \sqrt{2/\pi}\sin kx$ and we have what is called, to within a multiplicative constant, the ***Fourier sine transform***. If we define

$$\hat{f}_s(k) = \int_0^\infty f(x)\phi(x,k)\,\mathrm{d}x,$$

the inversion formula is

$$f(x) = \int_0^\infty \hat{f}_s(k)\phi(x,k)\,\mathrm{d}k.$$

In practice, most people work with $\sqrt{\pi/2}\,\hat{f}_s(k)$ and retain the factor $2/\pi$ in the inversion; the same remark about normalisation applies to the Fourier transform.

The lesson to be learnt from this discussion is that, if we define $\hat{f}_\lambda$ by (5.81), we should always be able to use the inversion formula

$$f(\mathbf{x}) = \int \hat{f}_\lambda \phi_\lambda\,\mathrm{d}\lambda, \tag{5.83}$$

as long as we are prepared to use results like (5.82) to define the normalisation. We must also be careful in defining the range of integration, which is very problem dependent and leads to a sum over a discrete set of λ when the spectrum is discrete. In such a case, $\hat{f}_\lambda$ is a superposition of delta functions and (5.83) is a generalised Fourier series.

We finally remind readers of the symmetry of (5.81) and (5.83) compared to the Fourier transform formulæ $\hat{f}(k) = \int f(x)\mathrm{e}^{ikx}\,\mathrm{d}x$ and $f(x) = (1/2\pi)\int \hat{f}(k)\mathrm{e}^{-ikx}\,\mathrm{d}k$. This reflects the fact that (5.81) depends on ϕ_λ being real. Since we use the complex eigenfunctions e^{ikx} for the Fourier transform, all inner products for such problems need to be defined for Hermitian operators, as explained in §4.4. Thus we need to define

$$(u,v) = \int_D u(\mathbf{x})\bar{v}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{5.84}$$

which means that, in (5.81), the Fourier coefficients are with respect to the conjugate eigenfunctions. When the eigenfunctions are complex, the crucial result on which the theory of transforms depends, and of which (5.82) is a special case, is

$$\int \phi_\lambda(\mathbf{x})\bar{\phi}_\lambda(\mathbf{x}')\,\mathrm{d}\lambda = \delta(\mathbf{x}-\mathbf{x}'), \tag{5.85}$$

where the integration is taken over whatever range is used in (5.81).

5.8 Transform solutions of elliptic problems

As happens all too often with very general theories, we soon encounter hideous technical complications when we try to apply the ideas of §5.7 to specific elliptic problems, even in two dimensions. This is simply because of the difficulty of identifying and characterising the eigenvalues and eigenfunctions of partial differential operators in any useful way. Fortunately, there is a widely occurring class of problems where we can decompose these eigenfunctions into those of ordinary differential operators, and we will henceforth restrict our attention to this class. They are characterised by the property that the variables are separable, so that the partial differential operator can be manipulated into the form $\mathcal{L}(x,y) = \mathcal{L}_1(x) + \mathcal{L}_2(y)$, where $\mathcal{L}_1$ and $\mathcal{L}_2$ only involve coefficients and differentials in x and y, respectively. For simplicity, we will also restrict attention to cases where the original differential equation is homogeneous, i.e. has zero right-hand side; when the right-hand side is non-zero it is possible to proceed by expanding it in terms of the eigenfunctions of $\mathcal{L}$, but this often leads to considerable technical difficulties. Assuming separability and homogeneity, the representation of u as a double series $u = \sum c_{mn}\phi_{mn}(x,y)$, where $\mathcal{L}\phi_{mn} = \lambda_{mn}\phi_{mn}$, simplifies to $u = \sum c_\lambda \phi_\lambda(x)\psi_\lambda(y)$, where $\mathcal{L}_1\phi_\lambda(x) = \lambda\phi_\lambda(x)$ and $\mathcal{L}_2\psi_\lambda(y) = -\lambda\psi_\lambda(y)$. For homogeneous problems in which separability is not possible, all that we can say is that the solution *can* be represented as a series in ϕ_{mn}, but an explicit formula for ϕ_{mn} is rarely available.

Let us now turn to some examples.

5.8.1 Laplace's equation with cylindrical symmetry: Hankel transforms

Suppose we wish to solve Laplace's equation in cylindrical polars,

$$\nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0, \tag{5.86}$$

in $z > 0$ with $u(r,\theta,0) = g(r)$, so that the solution is independent of θ, and with whatever decay rate is needed at infinity to ensure uniqueness. We recall that, to find the solution (5.60) of the half-plane problem in two dimensions, it was easy to take a Fourier transform in x and solve the resulting ordinary differential equations in y. This was because the eigenfunctions of $\mathrm{d}^2/\mathrm{d}x^2$ are so easy to find. When we separate the variables here we find

$$\frac{\mathrm{d}^2\psi_\lambda}{\mathrm{d}z^2} = -\lambda\psi_\lambda, \qquad \frac{\mathrm{d}^2\phi_\lambda}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\phi_\lambda}{\mathrm{d}r} = \lambda\phi_\lambda.$$

Hence we have no problem with the z derivatives in (5.86), which lead to eigenfunctions $\mathrm{e}^{-\sqrt{-\lambda}z}$, but to take a transform in r we need to consider the spectrum of $\mathrm{d}^2/\mathrm{d}r^2 + (1/r)\,\mathrm{d}/\mathrm{d}r$, which is not a self-adjoint operator. Hence, as in §4.5.2, for the function $\phi_\lambda(r)$ that describes the dependence of u on r, we consider not the eigenvalue problem

$$\frac{\mathrm{d}^2\phi_\lambda}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\phi_\lambda}{\mathrm{d}r} = \lambda\phi_\lambda$$

but, instead, the self-adjoint problem

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}\phi_\lambda}{\mathrm{d}r}\right) = \lambda r\phi_\lambda. \tag{5.87}$$

The factor r on the right-hand side introduces a slight complication into the argument in §5.7, namely that a weight function has to be introduced into the inner products: with reference to (5.81), if $\mathcal{L}$ is self-adjoint and $\mathcal{L}\phi_\lambda = \lambda q(\mathbf{x})\phi_\lambda$, where ϕ_λ is real, we simply find that, in one dimension, our basic Fourier transform formula must be changed to

$$\hat{f}_\lambda = \int f(x)q(x)\phi_\lambda(x)\,\mathrm{d}x, \quad f(x) = \int \hat{f}_\lambda\phi_\lambda(x)\,\mathrm{d}\lambda, \tag{5.88}$$

assuming now that ϕ_λ are normalised so that $\int q\phi_\lambda^2\,\mathrm{d}x = 1$. In terms of generalised functions, this equation merely reflects the fact that (5.85) has become

$$\int \phi_\lambda(x)\phi_\lambda(x')q(x')\,\mathrm{d}\lambda = \delta(x' - x). \tag{5.89}$$

For (5.87), the eigenfunctions are $\phi = J_0(kr)$ and $Y_0(kr)$, where $\lambda = -k^2$, $0 < k < \infty$, and J_0 and Y_0 are Bessel functions of zero order. The latter is discarded because it is unbounded as $r \to 0$ and we require the solution of (5.86) in the region $r \geqslant 0$. To conform with conventional usage, we label the eigenfunctions with $k = \sqrt{-\lambda}$ instead of λ and, as in §4.5.2, write the Hankel transform

$$\hat{u}(k,z) = \int_0^\infty rJ_0(kr)u(r,z)\,\mathrm{d}r; \tag{5.90}$$

we thereby obtain[80]

$$\frac{\mathrm{d}^2\hat{u}}{\mathrm{d}z^2} - k^2\hat{u} = 0 \quad \text{for } r > 0, \tag{5.91}$$

with

$$\hat{u}(k,0) = \hat{g}(k).$$

In order to write down the inversion formula for (5.90) from the recipe in §5.7.2, we must first normalise $J_0(kr)$. We have effectively done this in (4.48), but a more direct procedure is to revert to a limit of the type considered in §4.4 and consider functions $v(r)$ defined on $0 < r < R$, with $v(0)$ bounded and $v(R) = 0$, subsequently letting $R \to \infty$. The 'Fourier series' corresponding to the operator $\mathrm{d}/\mathrm{d}r(r\,\mathrm{d}/\mathrm{d}r)$ is

$$v(r) = \sum a_n J_0(k_n r), \tag{5.92}$$

where k_n is such that $J_0(k_n R) = 0$ and

$$a_n = \int_0^R rv(r)J_0(k_n r)\,\mathrm{d}r \bigg/ \int_0^R rJ_0^2(k_n r)\,\mathrm{d}r\,.$$

Now, letting $R \to \infty$ and using the famous result that $J_0(x)$ is approximately $\sqrt{2/\pi x}\cos(x - \pi/4)$ for large x, we find that $(1/R)\int_0^R rJ_0^2(k_n r)\,\mathrm{d}r \to 1/\pi k_n$.

[80] If you prefer not to write this down at once from §5.7, simply evaluate

$$\int_0^\infty \left\{J_0(kr)\left(\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + r\frac{\partial^2 u}{\partial z^2}\right) - u\left(\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}J_0(kr)\right) + k^2 rJ_0(kr)\right)\right\}\mathrm{d}r = 0.$$

Thus, since, for large k, the k_n are almost separated by π/R, the sum (5.92) can be thought of as a Riemann integral, giving

$$v(r) \to \int_0^\infty k\hat{v}(k)J_0(kr)\,\mathrm{d}k, \tag{5.93}$$

as $R \to \infty$, where

$$\hat{v}(k) = \int_0^\infty rv(r)J_0(kr)\,\mathrm{d}r; \tag{5.94}$$

this is just the formula before (4.48). The asymptotic behaviour of $J_0(x)$ for large x means that no normalising factor needs to be introduced into (5.93), in the same way that none would have been needed for Fourier transforms had we worked with $(1/\sqrt{2\pi})\mathrm{e}^{ikx}$ instead of e^{ikx}. Thus we can assert that[81]

$$u(r,z) = \int_0^\infty kJ_0(kr)\hat{u}(k,z)\,\mathrm{d}k$$

and we can now solve our problem by substituting $\hat{u} = \hat{g}(k)\mathrm{e}^{-kz}$ into this inversion formula.

Reverting to our discussion at the beginning of this section, we could have adopted a full-blooded approach and tackled the problem in r and z simultaneously as a brute-force application of the ideas of §5.7. The eigenfunctions would have been $\phi(r,z;k_1,k_2) = \mathrm{e}^{-k_1 z}J_0(k_2 r)$, with k_1 and k_2 real and positive, and writing

$$\hat{\phi} = \int_0^\infty \int_0^\infty \phi(r,z;k_1,k_2)u(r,z)\,r\,\mathrm{d}r\,\mathrm{d}z$$

would have given $(k_1^2 - k_2^2)\hat{\phi} = 0$. The only possibility would then be for $\hat{\phi}$ to be a generalised function[82] with support at $k_1 = k_2$ and this would also have led, eventually, to the same result.

We also point out that there is another important method for many practical problems for Laplace's equation with axial symmetry. In (5.63), we superimposed two-dimensional Green's functions, but for (5.86) we can superimpose three-dimensional ones to write the general solution as

$$\phi(r,z) = \int_{-\infty}^\infty \frac{g(\xi)\,\mathrm{d}\xi}{\sqrt{r^2+(z-\xi)^2}} \tag{5.95}$$

for some suitable weight function g (we also note that this would be reminiscent of the general solution (4.61) for the wave equation, were we to boldly replace $a_0 t$ by iz). It is fairly easy to see (as in Exercise 5.33) that ϕ decays at infinity and that $\phi = -2g(z)\log r + O(1)$ as $r \to 0$.

[81] The equation (5.89) reads

$$\int_0^\infty r'kJ_0(kr')J_0(kr)\,\mathrm{d}k = \delta(r'-r);$$

anyone puzzled by this can see the explanation about normalisation after (5.98).

[82] The interested reader should verify that the only distribution $f(x)$ that satisfies $xf(x) = 0$ for all x is a constant multiple of the delta function.

5.8.2 Laplace's equation in a wedge geometry; the Mellin transform

When the separation of variables strategy is applied to, say, Laplace's equation in a wedge with data on $\theta = 0$ and $\theta = \alpha$ in polar coordinates, it leads us to the eigenvalue problem

$$\frac{\mathrm{d}^2\phi_\lambda}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\phi_\lambda}{\mathrm{d}r} = \frac{\lambda}{r^2}\phi_\lambda$$

for the function ϕ_λ describing the r dependence of u, because the only term involving a θ derivative is $(1/r^2)\,\partial^2/\partial\theta^2$. As in (5.87), we write this in the self-adjoint form

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}\phi_\lambda}{\mathrm{d}r}\right) = \frac{\lambda}{r}\phi_\lambda, \tag{5.96}$$

with eigenfunctions $r^{\pm k}$, where $\lambda = k^2$. Suppose we require u to be bounded as $r \to 0$ and as $r \to \infty$. It is tempting to define $\hat{u}(k,\theta) = \int_0^\infty r^{k-1}u(r,\theta)\,\mathrm{d}r$ in line with (5.88), but we could never define an inversion formula because of the difficulty of normalising r^k. This is an all-too-common situation in transform theory but it can be overcome by the idea introduced in §4.4 of 'complexifying' the transform as a function of k. In Chapter 4 we found that we could only obtain bounded eigenfunctions of $\mathrm{d}^2/\mathrm{d}x^2$ as $|x| \to \infty$ if λ is real and negative; here we only get boundedness as $r \to 0, \infty$ if λ in (5.96) is real and negative, i.e. if k is purely imaginary. Hence, we define the *Mellin transform* by

$$\hat{u}(k,\theta) = \int_0^\infty r^{k-1}u(r,\theta)\,\mathrm{d}r \tag{5.97}$$

for suitable complex values of k. Instead of an inversion involving an integral like $\int_0^\infty r^k\hat{u}(k,\theta)\,\mathrm{d}k$, we expect

$$u(r,\theta) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} n(k)r^{-k}\hat{u}(k,\theta)\,\mathrm{d}k; \tag{5.98}$$

here $n(k)$ is another annoying normalising factor and the sign change in $r^k \to r^{-k}$ comes about because the fact that k is imaginary requires us to work with conjugate inner products, as in Fourier transforms. We hope that by now the reader is comfortable enough with generalised functions to say that

$$\begin{aligned}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} (r')^{k-1}r^{-k}\,\mathrm{d}k &= \mathrm{i}\int_{-\infty}^{\infty} (r')^{-1}\mathrm{e}^{\mathrm{i}k\log(r'/r)}\,\mathrm{d}k \\ &= 2\pi\mathrm{i}(r')^{-1}\delta\left(\log\left(\frac{r'}{r}\right)\right) \\ &= 2\pi\mathrm{i}\delta(r'-r);\end{aligned}$$

the penultimate step comes from (5.89) and the discussion before (4.29), and the final one from the fact that, if $f(x)$ is monotone and $f(0) = 0$, then $\delta(f(x)) = |f'(0)|^{-1}\delta(x)$. Hence $n(k) = 1/2\pi\mathrm{i}$. We remark that, in the same way that Fourier transforms can be applied to functions that are not square-integrable by 'shifting'

k up and down in the complex plane, so can (5.98) be made more widely applicable by taking the inversion contour from $c - \mathrm{i}\infty$ to $c + \mathrm{i}\infty$ for some appropriate real constant c.

We can make our usual statement that, although the Mellin transform is usually defined as (5.97) with k replaced by s, it would really be more self-consistent to replace k by $\mathrm{i}s$ (the same remark applies to the Laplace transform).

As an example, suppose that

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0,$$

with

$$u = u_0(r) \quad \text{on } \theta = \alpha \qquad \text{and} \qquad \frac{\partial u}{\partial \theta} = 0 \quad \text{on } \theta = 0,\ 0 < r < \infty.$$

When we proceed as with the Hankel transform, we find that

$$\frac{\mathrm{d}^2 \hat{u}}{\mathrm{d}\theta^2} - k^2 \hat{u} = 0, \tag{5.99}$$

with

$$\hat{u} = \hat{u}_0(k) \quad \text{on } \theta = \alpha, \qquad \frac{\mathrm{d}\hat{u}}{\mathrm{d}\theta} = 0 \quad \text{on } \theta = 0,$$

and so

$$\hat{u}(k, \theta) = \hat{u}_0(k)\frac{\cos k\theta}{\cos k\alpha}.$$

Thus

$$u(r, \theta) = \frac{1}{2\pi \mathrm{i}}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} r^{-k}\hat{u}_0(k)\frac{\cos k\theta}{\cos k\alpha}\,\mathrm{d}k. \tag{5.100}$$

If, say,

$$u_0 = \begin{cases} 1, & 0 < r < 1, \\ 0, & 1 < r < \infty, \end{cases}$$

we need to take $\Re k > 0$ in (5.97), so that $\hat{u}_0 = 1/k$ and then, with $k = \mathrm{i}\kappa$,

$$u = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\cosh \kappa\theta}{\cosh \kappa\alpha}\sin(\kappa \log r)\frac{\mathrm{d}\kappa}{\kappa},$$

the first term coming from the residue at $k = 0$.

*5.8.3 Helmholtz' equation

For this section only, we replace the k in Helmholtz' equation by unity, in order to avoid a notational clash. In a half-plane, transform methods for Helmholtz' equation, now written $(\nabla^2 + 1)\,u = 0$, are cursed by the presence of branch points in the transform plane (which we still denote by k). For example, when Dirichlet data $u(x,0) = g(x)$ is given on $y = 0$, the simple formula (5.62) for Laplace's equation is replaced by

$$\hat{u}(k,y) = \hat{g}(k)e^{-y\sqrt{k^2-1}} \tag{5.101}$$

and great care has to be taken over the definition of $\sqrt{k^2-1}$ and the appropriate choice of inversion contour (remember we have to consider complex functions u). We will content ourselves with just one transform solution which will be useful later in this chapter and in Chapter 8. This concerns scattering of radiation by a half-line and is called the *Sommerfeld problem*.

As in §5.1.5, we consider a plane sound wave in which the incoming field is $u_{\text{inc}} = e^{-ir\cos(\theta+\alpha)}$, incident at an angle α on a rigid barrier along the negative x axis, on which u vanishes (see Fig. 5.5). We are led to the following boundary value problem for Helmholtz' equation in $-\pi < \theta < \pi$:

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + u = 0, \tag{5.102}$$

with

$$u \to 0 \quad \text{as } \theta \to -\pi, \pi,$$

and u satisfies an appropriate outgoing condition as $r \to \infty$. In the 'shadow' region $\pi - \alpha < \theta < \pi$, this is simply (5.75), namely

$$\frac{\partial u}{\partial r} - iu = o(r^{-1/2}). \tag{5.103}$$

However, in the directly illuminated region $-\pi + \alpha < \theta < \pi - \alpha$, (5.103) must be satisfied by $u - e^{-ir\cos(\theta+\alpha)}$, and in the reflected region $-\pi < \theta < -\pi + \alpha$, it must be satisfied by $u - e^{-ir\cos(\theta+\alpha)} + e^{-ir\cos(\theta-\alpha)}$.

As in earlier examples, we have an easy eigenvalue problem for the operator $\partial^2/\partial\theta^2$, but now the eigenvalue problem in r takes the self-adjoint form

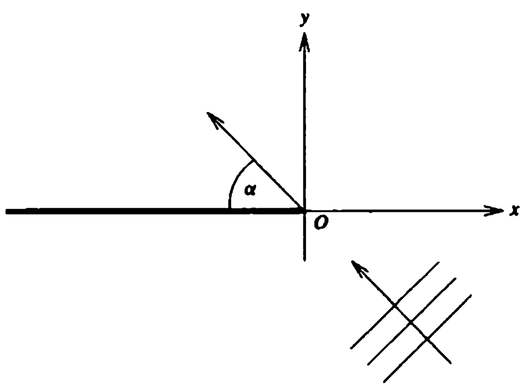


Fig. 5.5 The Sommerfeld problem.

$$\frac{d}{dr}\left(r\frac{d\phi_\lambda}{dr}\right) + r\phi_\lambda = \frac{\lambda\phi_\lambda}{r}. \tag{5.104}$$

Since the solutions of (5.104) behave like $r^{\pm\sqrt{\lambda}}$ or, when $\lambda = 0$, like $\log r$ as $r \to 0$, we require λ to be negative, as in the Mellin transform. Setting $\lambda = -\kappa^2$, we can then only satisfy the radiation condition (5.103) if ϕ_λ is proportional to

$$J_{i\kappa}(r) + iY_{i\kappa}(r) = H^{(1)}_{i\kappa}(r),$$

the Hankel function of complex order. Thus we define the *Kontorovich–Lebedev transform* of u as

$$\hat{u}_\kappa(\theta) = \int_0^\infty H^{(1)}_{i\kappa}(r)u(r,\theta)\frac{dr}{r},$$

and, in line with (5.91) and (5.99), we find that

$$\frac{d^2\hat{u}_\kappa}{d\theta^2} - \kappa^2\hat{u}_\kappa = 0. \tag{5.105}$$

Unfortunately, the generalisation of the argument in §5.7 now involves not only the intricacies of the normalisation, but also a very careful accounting in the Hermitian inner products. Hence we simply quote the result

$$u(r,\theta) = -\frac{1}{2}\int_{-\infty}^{\infty} \kappa\hat{u}_\kappa(\theta)J_{i\kappa}(r)\,d\kappa, \tag{5.106}$$

which can be manipulated into many other forms. Another justification of (5.106) is given in Exercise 5.14.

However, there is a further complication as far as the Sommerfeld problem is concerned. We have indicated the need to subtract out the incident field in order to be able to satisfy the condition (5.103), but this leaves us with a boundary condition for u on $\theta = \pm\pi$ which is non-zero as $r \to 0$. Thus, further contortions are needed before we can apply the Kontorovich–Lebedev transform, and we will not plough through to the final solution here; it is given in Exercise 5.39.

Not surprisingly, this solution can be derived by various other 'tricks', one of which is mentioned in the next section. However, the details are less important than the fact that even this hard example is readily amenable in principle to a systematic transform analysis.

We remark that among the many implications of the theory of elliptic equations for wave propagation in the frequency domain is the *principle of reciprocity*. This is the physical interpretation of the symmetry of Green's functions for self-adjoint elliptic problems. The statement from §5.6.2 that $G(\mathbf{x},\boldsymbol{\xi}) = G(\boldsymbol{\xi},\mathbf{x})$ for Helmholtz' equation says that the wave field at $\mathbf{x}$ caused by a 'point source' at $\boldsymbol{\xi}$ is identical to the wave field at $\boldsymbol{\xi}$ caused by a point source at $\mathbf{x}$. One very useful case is the limit $|\boldsymbol{\xi}| \to \infty$, when we can use the results of scattering problems, such as the Sommerfeld problem above, to infer the far-field radiated by a source near the scattering obstacle in an otherwise quiescent medium.

*5.8.4 Higher-order problems

We conclude this section with a brief discussion of a few points that may need to be borne in mind when Green's functions and transforms are applied to higher-order equations. The first concerns the biharmonic equation

$$\nabla^4 u = 0$$

for which the Green's function in two dimensions satisfies

$$\nabla^4 G = \delta(x-\xi)\delta(y-\eta). \tag{5.107}$$

One way to proceed in the whole of $\mathbb{R}^2$ is to invert the double Fourier transform $\left(k_1^4 + 2k_1^2k_2^2 + k_2^4\right)^{-1}$, but it is more straightforward to look for radially-symmetric solutions of

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)^2 G = 0 \quad \text{for } r > 0, \tag{5.108}$$

where $r^2 = (x-\xi)^2 + (y-\eta)^2$. Hence G is a linear combination of r^2, $r^2 \log r$, $\log r$ and a constant, and we might then guess that, since

$$\frac{1}{8\pi}\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)\left(r^2\log r - r^2\right) = \frac{1}{2\pi}\log r,$$

we might have

$$G = \frac{1}{8\pi}\left(r^2\log r - r^2\right). \tag{5.109}$$

However, to reassure ourselves that there should be no additional $\log r$ terms in G, we need to compute $\nabla^4 \log r$ in accordance with the procedure described in Exercises 4.13 and 4.21, where it was shown that $\nabla^2(\log r) = 2\pi\delta(x)\delta(y)$. We simply look at $(\nabla^4\psi, \log r) = \lim_{\epsilon\to 0}\iint_{r>\epsilon}\log r\,\nabla^4\psi\,\mathrm{d}x\,\mathrm{d}y$ for suitable test functions ψ. Since $\nabla^2\log r = 0$ when $r > 0$, repeated use of Green's theorem gives

$$\begin{aligned}(\psi, \nabla^4\log r) &= \lim_{\epsilon\to 0}\iint_{r>\epsilon}\psi\nabla^4\log r\,\mathrm{d}x\,\mathrm{d}y\\ &= \lim_{\epsilon\to 0}\left(\iint_{r>\epsilon}\nabla^4\psi\,\log r\,\mathrm{d}x\,\mathrm{d}y + \int_{r=\epsilon}\frac{\partial}{\partial r}\left(\nabla^2\psi\right)\log\epsilon\,\mathrm{d}s\right.\\ &\qquad\left. - \int_{r=\epsilon}\frac{1}{\epsilon}\nabla^2\psi\,\mathrm{d}s\right).\end{aligned}$$

The second term on the right-hand side tends to zero and the last to $-2\pi\nabla^2\psi(0,0)$ as $\epsilon \to 0$, which is a term we do not want, and hence there is no $\log r$ term in G.

The biharmonic equation may look to be a relatively benign generalisation of Laplace's equation. Indeed, because

$$\nabla^4 = 16\frac{\partial^4}{\partial z^2\,\partial\bar{z}^2},$$

its 'general solution' can be written down in two dimensions in terms of complex variables as $\Re\left(\bar{z}f(z) + g(z)\right)$, where $z = x + \mathrm{i}y$, $\bar{z} = x - \mathrm{i}y$ and f and g

are arbitrary analytic functions. However, it admits no maximum principle, and the fact that it is fourth order often destroys the possibility of easily finding explicit eigenfunction representations. Also, the fact that most practical, and probably well-posed, boundary value problems only have two pieces of data on the boundary of a closed domain D means that we can rarely expect to prove that $\iint_D u\nabla^4 v \,dx\,dy = \iint_D v\nabla^4 u \,dx\,dy$; hence, even if we have a complete set of eigenfunctions in which to expand, they are not usually orthonormal (see Exercise 5.28).

Another kind of complexity arises when we consider vector systems of equations. One that we have encountered in §4.7.2 concerns the steady case of Maxwell's equations in the form

$$\nabla \wedge \mathbf{H} = \mathbf{j}, \quad \nabla \cdot \mathbf{H} = 0. \tag{5.110}$$

It turns out that, in practice, these equations often have to be solved for $\mathbf{H}$ given both $\mathbf{j}$ and some physically relevant boundary conditions. Although we do not wish to present a systematic discussion of vectorial Green's functions until Chapter 9, we can point out a famous way of solving (5.110) using scalar Green's functions. We remove the constraint $\nabla \cdot \mathbf{H} = 0$ by writing $\mathbf{H} = \nabla \wedge \mathbf{A}$, where $\mathbf{A}$ is called the *magnetic vector potential*, and now we see that if we satisfy the 'gauge condition' $\nabla \cdot \mathbf{A} = 0$ we are simply left with a vector Poisson equation

$$\nabla^2 \mathbf{A} = -\mathbf{j}. \tag{5.111}$$

Our by-now-familiar Green's function argument gives the solution as

$$\mathbf{A}(\boldsymbol{\xi}) = \frac{1}{4\pi}\iiint \frac{\mathbf{j}(\mathbf{x})}{|\mathbf{x}-\boldsymbol{\xi}|}\,d\mathbf{x}$$

in $\mathbb{R}^3$, and, taking the curl, we find

$$\mathbf{H}(\boldsymbol{\xi}) = \frac{1}{4\pi}\iiint \frac{\mathbf{j}(\mathbf{x}) \wedge (\boldsymbol{\xi}-\mathbf{x})}{|\mathbf{x}-\boldsymbol{\xi}|^3}\,d\mathbf{x}, \tag{5.112}$$

which is the famous *Biot–Savart* law of electromagnetism. If $\mathbf{j}$ is simply a current in a wire, the calculation of $\mathbf{H}$ reduces to a single curvilinear integral.

5.9 Complex variable methods

At one level, it could be said that complex variable theory solves all problems for Laplace's equation in two dimensions, because the general solution is the real part of an analytic function of $x+iy$ (or $x-iy$). Indeed, we can formalise this statement by writing the solution of $\nabla^2 u = 0$ in $y > 0$, with $u(x,0) = f(x)$, as a Fourier inversion

$$u(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx-|k|y}\hat{f}(k)\,dk,$$

from which we find

$$u(x,y) = f_+(x-iy) + f_-(x+iy), \tag{5.113}$$

where

$$\hat{f}_+(k) = \begin{cases} \hat{f}(k), & k > 0, \\ 0, & k < 0, \end{cases} \qquad \hat{f}_-(k) = \begin{cases} 0, & k > 0, \\ \hat{f}(k), & k < 0. \end{cases}$$

It is interesting to compare this result with (4.67) for the wave equation; in the same vein, we can write the solution of the three-dimensional Laplace's equation in $z > 0$, with $u(x, y, 0) = f(x, y)$, as

$$u(x,y,z) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i(k_1 x + k_2 y) - kz} \hat{f}(k_1, k_2) \, dk_1 \, dk_2,$$

where $k^2 = k_1^2 + k_2^2$ and $k > 0$. Transforming to polar coordinates, we obtain

$$u(x,y,z) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-ik(x\cos\phi + y\sin\phi - iz)} \hat{F}(k, \phi) \, dk \, d\phi,$$

where $\hat{F} = 2\pi k \hat{f}$. Making the observation that led to (4.68), we derive the representation

$$u(x,y,z) = \int_0^{2\pi} F(x\cos\psi + y\sin\psi - iz, \psi) \, d\psi. \tag{5.114}$$

When complemented by a similar formula with z replaced by $-z$ (or, equivalently, one in which ϕ is shifted by π), we retrieve the general solution of Laplace's equation in three dimensions as a single integral. We also make the following remarks.

- If u is axisymmetric about the z axis, we can write

  $$F(x\cos\psi + y\sin\psi - iz, \psi) = \frac{1}{2} G(z + ir\cos(\theta - \psi)),$$

  where $x = r\cos\theta$ and $y = r\sin\theta$, and hence that

  $$u(x,y,z) = \int_0^{\pi} G(z + ir\cos\psi) \, d\psi; \tag{5.115}$$

  clearly $\pi G(z) = u(0, 0, z)$.
- We could relate these ideas to the three-dimensional generalisation of (5.95) on p. 189, namely

  $$u(x,y,z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{g(\xi, \eta) \, d\xi \, d\eta}{\sqrt{z^2 + (x-\xi)^2 + (y-\eta)^2}}.$$

  However, this representation involves convolution integrals with functions that are complex versions of those in (4.65). Here we simply mention that (5.95) is singular at the singularity distribution along $x = y = 0$. If we required an axisymmetric potential function that was analytic on the z axis, we could consider

  $$\int_{z-ir}^{z+ir} \frac{g(\xi) \, d\xi}{\sqrt{r^2 + (z-\xi)^2}},$$

  which would immediately retrieve (5.115) above. Note that this approach does not work for source distributions in two dimensions, such as (5.63) on p. 177. But, in two dimensions, we can simply resort to (5.113).

These observations are intriguing but, alas, they are usually of little help in solving practical boundary value problems. Instead, we must usually resort to the Green's function and transform techniques of the previous sections. Nevertheless, as we now explain, there are some powerful constructive complex variable methods that are especially valuable for Laplace's equation in two dimensions.

5.9.1 Conformal maps

Three of the most distinctive features of an analytic function $f(z)$ of a complex variable $z = x + iy$ are as follows.

1. Its real and imaginary parts are harmonic functions, i.e. they satisfy Laplace's equation in x and y.
2. As long as $f'(z) \neq 0$, $f(z)$ maps regions from the z plane into the plane of $\zeta = f(z)$ such that mapped curves intersect at an angle equal to that between the original curves.
3. Laplace's equation in $\mathbb{R}^2$ is 'conformally invariant', because when we write
$$\zeta = \xi + i\eta = f(z) = f(x + iy)$$
we find
$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = |f'(z)|^2 \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} \right).$$

These facts are of profound significance to the solution of Laplace's equation and its 'square', the biharmonic equation. We have already mentioned that, if a partial differential equation is invariant under certain classes of transformations, we may be able to perform some simplifications, but here the fact that the transformations involve functions of a complex variable makes things especially interesting. One consequence is that we can conformally map Laplace's equation problems into other Laplace's equation problems, although the boundary conditions may not be invariant under the map. However, one of the most sensational implications is that, if we have to solve Laplace's equation in a simply-connected closed domain D with prescribed Dirichlet, Robin or Neumann data, and we can explicitly map D conformally into the interior or exterior of the unit circle, or a half-plane, then we can use our knowledge of Green's functions and/or transforms to solve the mapped problem. The catch here is that we must find the conformal map explicitly; its existence is known from the Riemann mapping theorem,[83] although extra care must be taken when using this theorem for multiply-connected regions.

One immediate corollary is that Green's functions for Dirichlet and Neumann problems map into Green's functions. Considering the Dirichlet problem for simplicity, suppose $\zeta = f(z, z_0)$ maps the domain D into the unit circle in the ζ plane, with $z = z_0$ going into $\zeta = 0$, so that the ζ plane Green's function is $(1/2\pi)\log|\zeta|$. Then we claim
$$G(z, z_0) = \frac{1}{2\pi} \log|f(z, z_0)| \tag{5.116}$$

[83] We will say more about this in Chapter 7.

is the Green's function in the z plane. This function vanishes on ∂D and, as $z \to z_0$, $G \sim (1/2\pi)\log|f'(z_0, z_0)|\,|z - z_0|$, and so has the right behaviour since[84] $f'(z_0, z_0) \neq 0$. A trivial illustration is provided by

$$f(z, z_0) = \frac{a(z - z_0)}{a^2 - z\bar{z}_0}, \tag{5.117}$$

with a real, which takes $z = z_0$ onto $\zeta = 0$. Also, since $|a^2 - \bar{z}_0 z| = a|z - z_0|$ when $|z| = a$, this map takes $|z| = a$ onto $|\zeta| = 1$ and clearly agrees with (5.58). Equally, (5.59) is simply a deduction from the Cauchy integral formula (see Exercise 5.22).

Another simple application concerns the Green's function for the Dirichlet problem in a wedge $0 < \theta < \alpha$ in polar coordinates. Since $\zeta = z^{\pi/\alpha}$ maps this wedge into a half-plane, the formula before (5.60) gives

$$G(z, z_0) = \frac{1}{2\pi}\log\left|\frac{z^{\pi/\alpha} - z_0^{\pi/\alpha}}{z^{\pi/\alpha} - \bar{z}_0^{\pi/\alpha}}\right|; \tag{5.118}$$

this result can be obtained, eventually, by the method of images if α is a rational multiple of π, in which case the images terminate.[85]

We cite one very useful example of the direct use of conformal maps in aerodynamics [27]. Consider a streaming flow at angle α past an ellipse $x^2/a^2 + y^2/b^2 = 1$. The velocity potential ϕ satisfies $\nabla^2\phi = 0$, with $\partial\phi/\partial n = 0$ on the ellipse and

$$\phi = U_\infty(x\cos\alpha + y\sin\alpha) + o(r) \tag{5.119}$$

as $x^2 + y^2 \to \infty$, but we must remember to be careful about uniqueness since we are in an exterior domain. It is easy to look up in a book, or, better, use the fact that $x = a\cos\theta$ and $y = b\sin\theta$ on the ellipse, that the ellipse is given in terms of elliptic coordinates by $\xi = \xi_0$, $0 \leqslant \eta < 2\pi$, where

$$z = x + \mathrm{i}y = c\cosh(\xi + \mathrm{i}\eta) = c\cosh\zeta, \tag{5.120}$$

with $c^2 = a^2 - b^2$ and $\xi_0 = \frac{1}{2}\log\big((a + b)/(a - b)\big)$. Hence we simply need to find a function $\phi + \mathrm{i}\psi$ of ζ that is analytic in $\xi > \xi_0$, maps this region onto the exterior of the ellipse, and has zero imaginary part on $\xi = \xi_0$ (so that $\partial\phi/\partial n = 0$) and tends to $cU_\infty e^{-\mathrm{i}\alpha}\cosh\zeta$ as $\xi \to +\infty$ (to satisfy (5.119)). An answer is

$$\phi + \mathrm{i}\psi = A\cosh(\zeta - \xi_0 - \mathrm{i}\alpha), \tag{5.121}$$

where A is real and such that $Ae^{-\xi_0 - \mathrm{i}\alpha} = cU_\infty e^{-\mathrm{i}\alpha}$. Thus, with $A = cU_\infty e^{\xi_0}$, (5.121) with (5.120) appears to solve the problem. We will return to this later.

[84]The requirement that $\zeta = f(z)$ is one-to-one is tacit. This means that f is a *univalent* function of z, which is a stronger requirement than simply saying 'f' exists and is nowhere zero or infinite'.

[85]The question of the commensurability of wedge angles with π does not seem to matter so much for Laplace's equation as for Helmholtz' equation, which is not conformally invariant. But images can still sometimes work in the latter case.

*5.9.2 Riemann–Hilbert problems

Complex variable theory gives us the possibility of solving much more general boundary value problems for Laplace's equation than the Dirichlet and Neumann problems we have considered so far. In particular, much is known about how to extend analytic functions (and hence harmonic functions) across boundaries. This suggests that, instead of solving Laplace's equation on one side of a boundary with, say, Dirichlet data prescribed on that boundary, we may be able to solve it on *both* sides of the boundary, with a given relationship between the data on either side of the boundary. The crack and aerofoil problems of §5.1.6 are archetypal examples of this situation. Not only will we find that we can handle such problems successfully, but also, as an added bonus, we will be able to use our technique in the transform plane to solve some mixed boundary value problems, where the data switches from Neumann to Dirichlet.

The most stringent form of continuation of analytic functions is that of analytic continuation. Even though we know this is an ill-posed procedure, there are some situations where analytic continuation is useful, in particular when it can be carried out explicitly. One such case is when $\Im w(z) = 0$ on $y = 0$; then the *Schwarz reflection formula*

$$\overline{w(\bar{z})} = u(x, -y) - iv(x, -y)$$

gives the analytic continuation of $w(z) = u(x, y) + iv(x, y)$ from $y > 0$ into $y < 0$. This formula is a limiting case of the result that $\overline{w(a^2/\bar{z})}$ gives the analytic continuation into $|z| < a$ of a function $w(z)$ whose imaginary part vanishes on $|z| = a$. This latter result is of some value in solving homogeneous Neumann problems in irrotational flow (see Exercise 5.26).

In general, though, the singularities that inevitably occur when we continue analytically indicate that we have demanded too much regularity at the boundary. The motivation for *Riemann–Hilbert problems* comes from considering functions that are harmonic in a domain D excluding a curve Γ, across which they have a jump discontinuity of some kind rather than being analytic there; Γ may be a closed curve dividing D into two parts $D_{\pm}$, or it may be open. For example, if Γ is the real axis, we know from (5.60) that the functions

$$u_{\pm}(x, y) = \pm\frac{y}{\pi}\int_{-\infty}^{\infty}\frac{g(\xi)\,d\xi}{(x-\xi)^2 + y^2} \tag{5.122}$$

are harmonic in D_+ $(y > 0)$ and D_- $(y < 0)$, and satisfy the same Dirichlet data $u = g(x)$ on $y = 0$; however, $\partial u/\partial y$ suffers a jump across $y = 0$ (see Exercise 5.32). Similarly, for the Neumann problem, (5.63) shows that $\partial u/\partial y$ is continuous but u suffers a jump. We emphasise that we could only find a function that satisfies both Dirichlet and Neumann data on $y = 0$ by the ill-posed process of analytic continuation, but we can solve well-posed Dirichlet or Neumann problems in $y > 0$ or $y < 0$ at the expense of introducing singularities on the x axis.

This discussion paves the way for solving problems such as the thin aerofoil model (5.29) and (5.30) by asking more generally about the Riemann–Hilbert problem in which an analytic function $w(z)$ must be found in D in such a way that a linear combination of the limiting values of w as the curve Γ is approached from either side is prescribed. If Γ is closed, we write $w(z)$ as $w_+(z)$ or $w_-(z)$, as appropriate; otherwise, we can write $w(z)$ unambiguously in all of D except on Γ. The key lies in the famous *Plemelj formulæ* in which w is represented as a Cauchy integral

$$w(z) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(s)\,\mathrm{d}s}{s - z} \tag{5.123}$$

for some suitable function $f(s)$. A cheap way to see what happens as z approaches a point t at which Γ is smooth is to deform Γ as in Fig. 5.6 and take the limit $\epsilon \to 0$; note that this procedure requires that we define $f(s)$ away from Γ, which may not necessarily be allowed, but in fact everything can be proved as long as f is merely Hölder continuous. First, we denote the ***principal value*** of the contour integral by

$$\mathrm{PV} \int_\Gamma \frac{f(s)\,\mathrm{d}s}{s - t} = \lim_{\epsilon \to 0} \int_{\Gamma \setminus \partial D_\epsilon} \frac{f(s)\,\mathrm{d}s}{s - t}, \tag{5.124}$$

where ∂D_ϵ is the dotted segment in Fig. 5.6, which is in accord with our earlier use of the term in (5.67). When Γ is closed, we then find that, at any point $z = t$ on Γ, the limiting values of $w(z)$ on either side of Γ are

$$w_\pm(t) = \pm\frac{1}{2} f(t) + \frac{1}{2\pi \mathrm{i}} \mathrm{PV} \int_\Gamma \frac{f(s)\,\mathrm{d}s}{s - t},$$

and these yield the Plemelj formulæ

$$w_+(t) - w_-(t) = f(t), \tag{5.125}$$

$$w_+(t) + w_-(t) = \frac{1}{\pi \mathrm{i}} \mathrm{PV} \int_\Gamma \frac{f(s)\,\mathrm{d}s}{s - t}. \tag{5.126}$$

These formulæ still apply when Γ has end-points or corners, and f can be allowed to have integrable singularities at such points of Γ. This fact is important in what follows, when we will be interested in solving equations like (5.126) for f given $w_+ + w_-$. Remember that special care must be taken to prescribe the behaviour of the solutions of elliptic equations near places where the boundary data is singular.

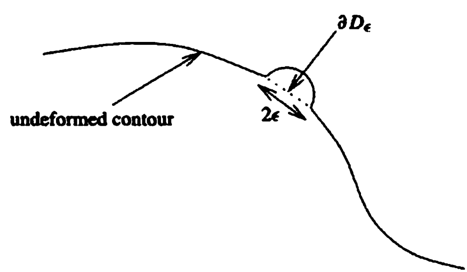


Fig. 5.6 Contour for the Plemelj formulæ.

Transcending these technicalities, what the Plemelj formulæ really tell us is that if, instead of prescribing the value of $w(z)$ on Γ, which is equivalent to prescribing two pieces of information for $\Re w$ and would lead to an ill-posed Cauchy problem for w, we prescribe the lesser information of $w_+ - w_-$ (or the values of $w_+ + w_-$ if we could solve the integral equation (5.126)), then we do have a seemingly well-posed problem for w. In fact, there are some boundary value problems for Laplace's equation where this allows us to write down the solution at once. For example, consider

$$\nabla^2 u = 0 \quad \text{in } D\text{, the region outside } \Gamma\text{: } y = 0,\ 0 < x < \mathrm{c}, \tag{5.127}$$

with

$$\lim_{y \downarrow \uparrow 0} u(x, y) = \pm g(x) \quad \text{for } 0 < x < \mathrm{c}, \tag{5.128}$$

respectively, and $u \to 0$ at infinity. Then $u = \Re w$, where $\Re(w_+ + w_-) = 0$ on Γ. Hence, from (5.126) we can find a solution in which f is real, and then from (5.125) $f = 2g$, so that

$$w(z) = \frac{1}{\pi \mathrm{i}} \int_0^c \frac{g(\xi)\,\mathrm{d}\xi}{\xi - z};$$

the real part of this equation is just an example of (5.122).

Equally, if, instead of (5.128), we have

$$\lim_{y \downarrow \uparrow 0} \frac{\partial u}{\partial y} = \pm g(x), \tag{5.129}$$

we consider $\mathrm{d}w/\mathrm{d}z = \partial u/\partial x - \mathrm{i}\,\partial u/\partial y$ to obtain $u = \Re w$, where (5.126) shows that there is a solution in which f is purely imaginary on the x axis. Then, from (5.125), $f = -2\mathrm{i}g$ there, and hence

$$\frac{\mathrm{d}w}{\mathrm{d}z} = -\frac{1}{\pi} \int_0^c \frac{g(\xi)\,\mathrm{d}\xi}{\xi - z}.$$

Integration gives that

$$u(x, y) = \frac{1}{2\pi} \int_0^c f(\xi) \log\left((x - \xi)^2 + y^2\right) \mathrm{d}\xi \tag{5.130}$$

to within an additive constant, as in (5.63). Note that both (5.122) and (5.130) can be interpreted as distributions of singularities along $y = 0$, $0 < x < \mathrm{c}$, or equivalently as Green's function formulæ. Physically, the singularities in (5.122) are called dipoles because they involve derivatives of the 'source' or 'monopole' $(1/2\pi)\log r$ that appears in (5.130).

We cannot use the Plemelj formulæ directly to solve boundary value problems for Laplace's equation unless the values of the function or its normal derivative are precisely equal and opposite. This is not the case, say, for the aerofoil problem (5.29) and (5.30), but there is a trick we can always use to turn arbitrary Riemann problems into ones we can solve directly. Suppose all we know is that

$$\alpha(t)w_+(t) + \beta(t)w_-(t) = \gamma(t) \tag{5.131}$$

on Γ, with α, β and γ prescribed and analytic, and α and β non-zero. We look first at the homogeneous problem

$$\alpha(t)W_+(t) + \beta(t)W_-(t) = 0, \tag{5.132}$$

which we can solve in principle since

$$\log W_+ - \log W_- = \log\left(\frac{-\beta}{\alpha}\right); \tag{5.133}$$

hence, from the Plemelj formulæ for $\log W$,

$$W_\pm(z) = \exp\left(\frac{1}{2\pi \mathrm{i}}\int_\Gamma \frac{\log(-\beta/\alpha)\,\mathrm{d}\xi}{\xi - z}\right).$$

At this point we are reminded of our earlier strictures about non-uniqueness, because different choices for the branch of the logarithm give rise to different W; clearly, any entire function can be added to $\log W$ without affecting (5.133). Note also that, if D extends to infinity, our solution for W automatically satisfies $\log W = O(1/z)$ there. Setting this aside for the moment, we go back to (5.131), which becomes

$$\left(\frac{w}{W}\right)_+ - \left(\frac{w}{W}\right)_- = \frac{\gamma}{\alpha W_+} = -\frac{\gamma}{\beta W_-},$$

and which can be solved for w/W, and hence w, explicitly. The resulting formula looks more complicated than it really is, but it can contain the answers to some key technical questions, as illustrated by the aerofoil problem. We only consider the most interesting case of a vanishingly-thin wing in which $f_+ = f_- = f$, so that, from (5.30),

$$\frac{\partial\tilde{\phi}}{\partial y} \to Uf'$$

as y approaches Γ, the line $y = 0$, $0 < x < c$, from above and below. We work with $w(z)$, where $\mathrm{d}w/\mathrm{d}z = \partial\tilde{\phi}/\partial x - \mathrm{i}\,\partial\tilde{\phi}/\partial y$ and

$$\Im\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_+ = \Im\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_- = -Uf'. \tag{5.134}$$

Hence, $(\mathrm{d}w/\mathrm{d}z)_+ - (\mathrm{d}w/\mathrm{d}z)_-$ is real and we write

$$\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_+ - \left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_- = g(x), \tag{5.135}$$

where, from (5.123) and (5.125),

$$\frac{\mathrm{d}w}{\mathrm{d}z} = \frac{1}{2\pi \mathrm{i}} \int_0^c \frac{g(\xi)\,\mathrm{d}\xi}{\xi - z}. \tag{5.136}$$

Now the other Plemelj formula (5.126) gives that

$$\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_+ + \left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_- = \frac{1}{\pi \mathrm{i}} \mathrm{PV} \int_0^c \frac{g(\xi)\,\mathrm{d}\xi}{\xi - x}$$

is purely imaginary, but all that (5.134) tells us is that

$$2Uf' = \frac{1}{\pi} \mathrm{PV} \int_0^c \frac{g(\xi)\,\mathrm{d}\xi}{\xi - x}. \tag{5.137}$$

The Plemelj formulæ have merely turned the problem into a *Cauchy singular integral equation* for g.

To proceed further, we use (5.132) to turn what we know about $(\mathrm{d}w/\mathrm{d}z)_+ + (\mathrm{d}w/\mathrm{d}z)_-$ into an equation for $(\mathrm{d}w/\mathrm{d}z)_+ - (\mathrm{d}w/\mathrm{d}z)_-$, so we let W be such that

$$W_+(x) + W_-(x) = 0. \tag{5.138}$$

Then our desired formula for $\mathrm{d}w/\mathrm{d}z$ comes from

$$\frac{1}{W_+}\left[\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_+ + \left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)_-\right] = -\frac{2\mathrm{i}Uf'}{W_+}; \tag{5.139}$$

the left-hand side is $((1/W)\,\mathrm{d}w/\mathrm{d}z)_+ - ((1/W)\,\mathrm{d}w/\mathrm{d}z)_-$ and hence, from (5.125) and (5.123),

$$\frac{\mathrm{d}w}{\mathrm{d}z} = -\frac{W(z)}{\pi} \int_0^c \frac{Uf'(\xi)\,\mathrm{d}\xi}{W_+(\xi)(\xi - z)}. \tag{5.140}$$

Thus, from (5.135), we find that

$$g(x) = \frac{2W_+(x)}{\pi} \mathrm{PV} \int_0^c \frac{Uf'(\xi)\,\mathrm{d}\xi}{W_+(\xi)(\xi - x)}, \tag{5.141}$$

which provides the solution of (5.137).[86]

But what is $W(z)$? It is possible to solve (5.138) by inspection, but we can be more systematic and take logarithms to give

$$\log W_+ - \log W_- = (2k+1)\mathrm{i}\pi, \quad k \text{ an integer}; \tag{5.142}$$

hence, by (5.125) and (5.123),

$$\log W = \frac{2k+1}{2} \int_0^c \frac{\mathrm{d}\xi}{\xi - z} = \frac{2k+1}{2} \log\left(\frac{z - c}{z}\right).$$

Thus

$$W_+(x) = \text{constant} \cdot \left(\frac{c - x}{x}\right)^{(2k+1)/2}. \tag{5.143}$$

It is implicit in (5.143) that the branch cut for $((c - z)/z)^{(2k+1)/2}$ is taken between $z = 0$ and $z = c$, which is necessary because W is analytic at $z = \infty$.

[86] Although (5.140) and (5.136) look different, they are the same when g satisfies (5.141).

Now, at last, we must address the uniqueness question. As anticipated from many earlier remarks, we expect further information about the behaviour of the solution at the leading and trailing edges of the wing to be necessary for uniqueness,[87] and this is true with a vengeance here. It would take us too far into aerodynamics to explain properly why the physically relevant solution should have $\mathrm{d}w/\mathrm{d}z$ bounded near $z = c$ and tending to infinity like $z^{-1/2}$ near $z = 0$, a requirement known as the *Kutta–Joukowsky condition*. However, if we assume that this is what is required, and retain the condition $\mathrm{d}w/\mathrm{d}z = O(1/z)$ at infinity, we simply need to take $k = 0$ in (5.143), the branch cut for W being along the aerofoil. For the record, our final formula (5.140) is

$$\frac{\mathrm{d}w}{\mathrm{d}z} = -\frac{\mathrm{i}U}{2\pi}\left(\frac{z-\mathrm{c}}{z}\right)^{1/2}\int_0^{\mathrm{c}}\left(\frac{\xi}{\mathrm{c}-\xi}\right)^{1/2}\frac{f'(\xi)\,\mathrm{d}\xi}{\xi - z}. \tag{5.144}$$

An interesting piece of quality control is possible to check this result. A simple aerofoil shape for which it is possible to obtain the streaming flow exactly is the ellipse, for which a putative solution is (5.121). However, if we let $b \to 0$ in (5.121) so that the ellipse becomes a flat plate, we soon find that we do *not* obtain agreement with (5.144) as α, which is effectively $f'(\xi)$, tends to zero; indeed, when $|z| \to \infty$, (5.144) gives $\mathrm{d}w/\mathrm{d}z = \mathrm{i}\Gamma U/2\pi z + o(1/z)$, where $\Gamma = \int_0^{\mathrm{c}} (\xi/(\mathrm{c}-\xi))^{1/2} f'(\xi)\,\mathrm{d}\xi$, but (5.121), in which w is now the difference between the complex potential and $U_\infty \mathrm{e}^{-\mathrm{i}\alpha} z$, gives that $\mathrm{d}w/\mathrm{d}z$ is of $O(1/z^2)$. This is because in (5.121) we omitted to include any eigensolutions[88] in which w is of the form (real constant)$\cdot\mathrm{i}\log z + O(1)$ as $z \to \infty$. The simplest way to do this is to map the ellipse to a circle and add an eigensolution $\phi = K\theta$ in polar coordinates (see Exercise 5.27). Now this formula only agrees with (5.144) when we choose the particular value of K, namely $\Gamma U/2\pi$, that makes $\mathrm{d}w/\mathrm{d}z$ finite at the trailing edge $z = \mathrm{c}$; putting $K = 0$ would correspond to setting $W = (z(z-\mathrm{c}))^{-1/2}$ after (5.142). Moreover, by considering the pressure in the fluid, it can be shown that the lift on the aerofoil is directly proportional to K; hence the practical point as far as flight is concerned is that, without the eigensolution brought in by the Kutta–Joukowsky condition, the aerofoil cannot generate any lift!

The aerofoil problem is an important, but special, problem for which we are lucky that the Plemelj formulæ lead immediately to a physically acceptable solution, because (5.135) has the correct behaviour as $z \to \infty$. In the examples which follow, the interplay between the behaviour at infinity and the singularities at the ends of the interval is less straightforward.

*5.9.3 Mixed boundary value problems and singular integral equations

We have encountered a prototypical mixed boundary value problem in our model (5.31) and (5.32) for fracture mechanics. In fact, it is one of the simplest elliptic

[87] Were we to relax the requirement that $\log W$ be of the form (5.124), we could obtain many more solutions, such as $W(z) = z^{(2m+1)/2}(z-\mathrm{c})^{(2n+1)/2}$, where m and n are integers.

[88] The importance of the double-connectedness of the flow region in admitting these eigensolutions has been anticipated at the end of §5.2.1.

problems with singularities in the boundary data, yet it can be used to illustrate some very complicated methodology. We recall that we have to solve $\nabla^2 w = 0$ in $\mathbb{R}^2$ with the slit $y = 0$, $-c < x < c$, (i.e. Γ) removed, with $\partial w/\partial y = 0$ on $y = 0$, $-c < x < c$, and $w = 0$ on the rest of the x axis, and with $w = \tau y + O(1)$ at infinity. We write $w = \tau y + u$ for simplicity to give

$$\nabla^2 u = 0 \quad \text{for } y > 0, \tag{5.145}$$

with

$$u = 0 \quad \text{on } y = 0, |x| > c, \qquad \frac{\partial u}{\partial y} = -\tau \quad \text{on } y = 0, |x| < c, \tag{5.146}$$

and, of course, we require not only that $u \to 0$ as $x^2 + y^2 \to \infty$, but also some specification of the behaviour of u at the crack tips $(\pm c, 0)$.

One way to proceed is to represent u as the response to a distribution of singularities[89] along the crack; following the calculation that stemmed from (5.66), we write

$$u(x,y) = \frac{1}{\pi}\int_{-c}^{c} g(\xi)\tan^{-1}\left(\frac{y}{\xi - x}\right)\mathrm{d}\xi \quad \text{for } -\frac{\pi}{2} < \tan^{-1}\left(\frac{y}{\xi - x}\right) < \frac{\pi}{2}, \tag{5.147}$$

so that, for $y \neq 0$,

$$\frac{\partial u}{\partial y} = \frac{1}{\pi}\int_{-c}^{c} \frac{g(\xi)(\xi - x)\,\mathrm{d}\xi}{(\xi - x)^2 + y^2},$$

and, as in (5.68), it is easy to show that we are left with the singular integral equation

$$-\tau = \mathrm{PV}\frac{1}{\pi}\int_{-c}^{c} \frac{g(\xi)\,\mathrm{d}\xi}{\xi - x} \quad \text{for } |x| < c. \tag{5.148}$$

From (5.137), this is just the non-trivial problem of a flat aerofoil. However, whereas the nature of the singularities at $z = -c$, c and ∞ in the aerofoil problem make it solvable by direct application of the Plemelj formulæ to (5.123), the physical requirements of the crack model mean that extra eigensolutions must be introduced. These eigensolutions can be realised by noticing that W, as introduced in (5.138), can be multiplied by $(z+c)^m(z-c)^n$, where m and n are integers which must be chosen to give the correct behaviour at $z = -c$, c and ∞, corresponding to the physical requirement that $|\nabla u|$ grows at most as the inverse square root of distance from the crack tips and $u \to 0$ at infinity. Fortunately, in this instance we can avoid solving a Riemann–Hilbert problem simply by spotting that the relevant solution is

$$u = \Im\left(-\tau z + \tau\sqrt{z^2 - c^2}\right), \tag{5.149}$$

where $\sqrt{z^2 - c^2} \to z$ as $z \to \infty$ (alternatively, see Exercise 5.29). This reveals the famous phenomenon of *stress intensification* at the tips: we find that, on $y = 0$,

[89]In solid mechanics, such singularities are sometimes referred to as (virtual) *dislocations*.

$$u = \begin{cases} \tau\sqrt{c^2 - x^2}, & |x| < c, \\ 0, & |x| > c, \end{cases}$$

while

$$\frac{\partial u}{\partial y} = \begin{cases} -\tau, & |x| < c, \\ \tau\left(x/\sqrt{x^2 - c^2} - 1\right), & |x| > c. \end{cases}$$

The size of the coefficient τ of the stress in $|x| > c$ can make all the difference between a brittle material failing or not! But the interesting mathematical result is that the relevant solution of (5.148) is[90] $g(x) = -\tau x/\sqrt{c^2 - x^2}$.

We note that both this problem and the aerofoil problem display a kind of 'rigidity' in respect of their singularity behaviour. In both cases, we could have made the behaviour at some of the singular points (i.e. the ends of Γ and infinity) better or worse at the expense of making that at the others worse or better. Equally we could, and indeed have, made the behaviour at the trailing edge of an aerofoil better than at the leading edge, but we could never attain finite velocity at both edges unless the flow were symmetric about the x axis.

*5.9.4 The Wiener–Hopf method

Our experience with aerofoil and fracture models now puts us in a position to describe the most famous systematic technique associated with mixed boundary value problems. This is the so-called *Wiener–Hopf method* and we do not even need a model as complicated as (5.145) and (5.146) to illustrate it. Suppose we have the elliptic mixed boundary value problem

$$\nabla^2 u = 0 \quad \text{for } y > 0,$$

with

$$u = 0 \quad \text{on } y = 0, x > 0, \qquad \frac{\partial u}{\partial y} = 0 \quad \text{on } y = 0, x < 0,$$

and, in polar coordinates, $u = r^{1/2}\sin(\theta/2) + O(r^{-1/2})$ as $r^2 = x^2 + y^2 \to \infty$, together, of course, with a prescription of the singularity at the origin to which we will return shortly. The answer is trivial to spot, but supposing we steadfastly persist with a Fourier transform approach, writing $\hat{u}(k, y) = \int_{-\infty}^{\infty} u(x, y)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x$, we obtain

$$\frac{\mathrm{d}^2\hat{u}}{\mathrm{d}y^2} - k^2\hat{u} = 0, \quad \text{so} \quad \hat{u} = A(k)\mathrm{e}^{-|k|y}.$$

We do not know $A(k)$ because we know neither u nor $\partial u/\partial y$ all the way along $y = 0$. But we could set

[90] You can check this result by writing $\xi = c\sin\theta$ in (5.148) to give

$$\mathrm{PV}\int_{-c}^{c} \frac{\xi}{\sqrt{c^2 - \xi^2}}\frac{\mathrm{d}\xi}{\xi - x} = \mathrm{PV}\int_{-\pi/2}^{\pi/2} \frac{\sin\theta\,\mathrm{d}\theta}{\sin\theta - x} = \pi + x\,\mathrm{PV}\int_{-1}^{1} \frac{2\,\mathrm{d}t}{2t - x(1 + t^2)},$$

where $t = \tan(\theta/2)$. The final integral can be evaluated in terms of elementary functions and found to be zero. An alternative derivation of the result using contour integration is given in Exercise 5.30.

$$\frac{\partial u}{\partial y}(x,0) = \begin{cases} f(x), & x > 0, \\ 0, & x < 0, \end{cases} \qquad u(x,0) = \begin{cases} 0, & x > 0, \\ g(x), & x < 0, \end{cases}$$

where $f(x)$ is unknown for $x > 0$ and, conversely, $g(x)$ is unknown for $x < 0$. We then obtain $A(k) = \hat{g}(k) = -\hat{f}(k)/\,|k|$. This gives just two equations for the three unknowns A, $\hat{f}$ and $\hat{g}$ but there are two vital pieces of information from which we can squeeze the solution. The first is that, since $\hat{f}(k) = \int_0^\infty f(x)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x$, the integral exists and $\hat{f}(k)$ is analytic if $\Im k$ is large enough and positive, specifically in an upper half-plane; the second is that, similarly, $\hat{g}(k)$ is analytic if $\Im k$ is large enough and negative, specifically in a lower half-plane.

Hence we have something very like a Riemann–Hilbert problem because there may be a line in the complex k plane, probably $\Im k =$ constant, on which $\hat{f}$ and $\hat{g}$ are both analytic, or there may even be a strip where the regions of analyticity overlap. If so, the traditional Wiener–Hopf procedure is to write $\hat{f} = \hat{f}_+$ and $\hat{g} = \hat{g}_-$ (the subscripts indicate the domain of analyticity), and recast

$$\hat{f}_+(k) + |k|\,\hat{g}_-(k) = 0$$

in the form

$$\hat{f}_+ + \frac{K_-}{K_+}\hat{g}_- = 0, \tag{5.150}$$

where the 'factors' $K_+(k)$ and $K_-(k)$ of $|k|$ are analytic in upper and lower half-planes of the complex k plane, respectively. If (5.150) holds, then the argument usually goes that $\hat{f}_+K_+$ is analytic in an upper half-plane and its analytic continuation into the lower half-plane is, by (5.150), equal to $-\hat{g}_-K_-$, which is known to be analytic there. Thus $\hat{f}_+K_+$, and hence $-\hat{g}_-K_-$, is entire, that is it has no singularities for finite k, and both these functions are equal to an entire function $\widetilde{E}(k)$, say. Determining this entire function is easy using Liouville's theorem as long as the behaviour of u at infinity is prescribed sufficiently carefully, but the 'factorisation' is the spectre that hangs over the whole procedure. In principle, it holds no terrors for people familiar with Riemann–Hilbert problems because, in order to write an arbitrary non-zero function $K(k)$ as a ratio of '$-$' and '$+$' functions K_-/K_+, we first write $\gamma(k) = -\log K(k)$ and consider

$$G(k) = \frac{1}{2\pi \mathrm{i}}\int \frac{\gamma(k')}{k'-k}\,\mathrm{d}k'$$

along a contour $\Im k = k_0$; then this defines one function, G_+, that is analytic in $\Im k > k_0$, and another one, G_-, analytic in $\Im k < k_0$. Moreover, by the Plemelj formula, the difference between these two functions is $\gamma(k)$ as $\Im k \to k_0 \pm 0$. We have thus identified two functions $G_\pm$ such that $\gamma(k) = G_+ - G_-$, and hence we have performed our 'factorisation' for the function $K(k) = \exp(-\gamma(k))$. Of course, the resulting formulæ are just as unpleasant as those encountered in most

Riemann–Hilbert problems. Indeed, although it is possible to carry out this procedure for $\log|k|$ and hence complete the calculation leading to (5.150) after suitable contortions with the integrals, we can fortunately spot that we can write

$$|k| = \lim_{\epsilon\to 0}(k - \mathrm{i}\epsilon)^{1/2}(k + \mathrm{i}\epsilon)^{1/2},$$

where the branch cuts must lie along $(\mathrm{i}\epsilon, \mathrm{i}\infty)$ and $(-\mathrm{i}\epsilon, -\mathrm{i}\infty)$, respectively. Hence, we have the possible factorisations

$$K_-(k) = (k - \mathrm{i}\epsilon)^{1/2}E(k), \quad \text{the branch cut going from } \mathrm{i}\epsilon \text{ to } +\mathrm{i}\infty,$$
$$K_+(k) = (k + \mathrm{i}\epsilon)^{-1/2}E(k), \quad \text{the branch cut going from } -\mathrm{i}\epsilon \text{ to } -\mathrm{i}\infty,$$

where $E(k)$ is some other entire function. So, letting $\epsilon \to 0$ and keeping the branch cuts as above, we have

$$k^{-1/2}\hat{f}_+(k) = -k^{1/2}\hat{g}_-(k) = \frac{\widetilde{E}(k)}{E(k)}.$$

Now we guess that the singular behaviour we need to specify as $r \to 0$ is $g(r) = O(r^{1/2})$ and $f(r) = O(r^{-1/2})$; it will soon become apparent that any other power-law behaviour in r would lead to non-existence or non-uniqueness. Then it is plausible (and this is shown in Exercise 5.35) that, as $k \to \infty$,

$$\hat{g}_-(k) = O\left(\int_{-\infty}^{0}(-x)^{1/2}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x\right) = O(k^{-3/2})$$

for $\Im k < 0$, and

$$\hat{f}_+(k) = O\left(\int_{0}^{\infty}x^{-1/2}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x\right) = O(k^{-1/2})$$

for $\Im k > 0$. Hence, as $k \to \infty$,

$$\begin{aligned} k^{-1/2}\hat{f}_+(k) &= -k^{1/2}\hat{g}_-(k) \\ &= \frac{\widetilde{E}(k)}{E(k)} \\ &= O\left(\frac{1}{k}\right). \end{aligned}$$

Now we are ready to use Liouville's theorem. Provided that $E(k)$ *only* vanishes at $k = 0$, we have that $\widetilde{E}(k)/E(k) = C/k$ for some constant C. In order for E to have this property, when we choose the factors K_+ and K_-, we must ensure that they do not vanish in the upper and lower half-planes, respectively. If we did not insist on this condition, we would not be able to pin down $\widetilde{E}/E$, and hence $\hat{f}_+(k)$ and $\hat{g}_-(k)$.

We now have $\hat{f}_+(k) = Ck^{-1/2}$, and so, by inversion,

$$f_+(x) = \frac{C}{2\pi}\int_{-\infty+i0}^{\infty+i0} \frac{e^{-ikx}\,dk}{\sqrt{k}} = \frac{iC}{2\pi\sqrt{-x}}\int_{-\infty+i0}^{\infty+i0} \frac{e^{-is}\,ds}{\sqrt{s}}.$$

Finally, with a suitable choice of C (which we could also have made by inspection from the behaviour of $\hat{f}(k)$ for small k, once it has been identified), we retrieve our expected result that

$$f = -\frac{1}{2r^{1/2}},$$

since, of course, $u \equiv r^{1/2}\sin(\theta/2)$.

The example above does scant justice to the Wiener–Hopf method, which is undoubtedly the most powerful practical method to solve many mixed boundary value problems, especially for Helmholtz' equation and its generalisations. As an illustration, we refer the reader to [24] where the solution of the Sommerfeld problem is derived.

We conclude our discussion with two final remarks about singularities in elliptic problems.

*5.9.5 Singularities and index

The questions raised in the introductory section concerning the difference between 'crack' problems and 'contact' problems in elasticity further highlight the importance of prescribing the singular behaviour of solutions of elliptic equations near the boundary. It would take too long to go into the practical modelling of contact or *Hertz problems* here because they involve configurations at least as complicated as plane strain (see §5.1.4.4) and hence the biharmonic equation. Suffice it to say that they involve mixed boundary value problems where the perimeter of the contact region marks the change in the boundary conditions and hence the location of the singularities. However, contact problems differ from many crack problems in that the contact region is not known *a priori*; it can decide for itself where it would like to be, and the idea is that the contact region should be determined by the fact that the strong stress intensification that we find in fracture does not occur. This class of problems will be considered more thoroughly in Chapter 7, but there is one configuration that illustrates the contrast between crack and contact problems and is simple enough to be considered here.

Suppose that, instead of considering contact between two general elastic bodies, we take a large smooth flat membrane initially in the (x, y) plane, and deform it by requiring it to lie above a thin smooth wire $z = f(x)$, lying in $y = 0$; we assume for simplicity that f is a smooth, even function with $d^2f/dx^2 < 0$ and $f(0) > 0$. Thus, the transverse displacement $z = u(x, y)$ satisfies

$$\nabla^2 u = 0 \quad \text{except on } y = 0,\ -c \leqslant x \leqslant c,$$

with

$$u(x,0) = f(x) \quad \text{for } |x| < c, \qquad \frac{\partial u}{\partial y}(x,0) = 0 \quad \text{for } |x| > c,$$

where the unknown number c defines the boundary of the contact region; we do ***not*** expect any stress intensification to occur at $x = \pm c$. We also need to prescribe the behaviour of u as $x^2+y^2 \to \infty$, and we recall that in the crack problem (5.145) and (5.146) we prescribed a tensile stress so that $|\nabla u| \to$ constant at large distances; here, we expect $|\nabla u|$ to be much smaller at infinity.[91]

Using our theory of mixed boundary value problems, we set $w(z) = u + \mathrm{i}v$ and find immediately that the solution that is symmetric in x, has no stress intensification at $(\pm c, 0)$, and is as small as possible as $|z| \to \infty$ is

$$\frac{\mathrm{d}w}{\mathrm{d}z} = \frac{\partial u}{\partial x} - \mathrm{i}\frac{\partial u}{\partial y} = \frac{\mathrm{i}}{\pi}\sqrt{c^2 - z^2}\int_{-c}^{c} \frac{1}{\sqrt{c^2-\xi^2}}\frac{\mathrm{d}f}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\xi - x}.$$

Note the contrast between this formula and a typical fracture formula, in which both the square roots would be reciprocated. Also, since

$$\int_{-c}^{c} \frac{1}{\sqrt{c^2-\xi^2}}\frac{\mathrm{d}f}{\mathrm{d}\xi}\,\mathrm{d}\xi = 0$$

by symmetry, u grows only logarithmically in $|z|$ as $|z| \to \infty$, so that, far from the obstacle, u is approximately proportional to the Green's function for Laplace's equation, centred at the obstacle; thus, at large distances, the effect of the obstacle is that of a point force.

Crucially, c must be such that the membrane is in equilibrium with the forces applied at its boundary. Writing $u = K\log(x^2+y^2)^{1/2} + O(1)$ at infinity, where K tells us how hard we are pushing the membrane down onto the obstacle, we see that

$$\frac{1}{\pi}\int_{-c}^{c} \frac{\xi}{\sqrt{c^2-\xi^2}}\frac{\mathrm{d}f}{\mathrm{d}\xi}\,\mathrm{d}\xi = K. \tag{5.151}$$

It is this condition that finally determines c.

The way in which the strengths of the singularities enter into these models from mechanics illustrates the general idea of characterising singularities in terms of an *index*. This concept has its basis simply in the computation of the change in the argument of an analytic function described by a Cauchy integral as we traverse a closed curve in the complex plane. This argument principle is equivalent to counting the number of zeros of the analytic function whose real part is the relevant

[91] We could equally consider the problem of closing a previously opened crack in which a displacement u_0 such that

$$u_0(x,0) = \begin{cases} f(x), & |x| < c_0, \\ 0, & |x| > c_0 \end{cases}$$

has been set up. Then the boundary conditions would be

$$\frac{\partial u}{\partial y}(x,0) = 0 \quad \text{for } |x| < c, \qquad u = \begin{cases} -f(x), & c < |x| < c_0, \\ 0, & c_0 < |x|. \end{cases}$$

It is now possible for $u = -\tau y + O(1)$ as $y \to \infty$, where τ is the crack closing stress, and our task is to find c such that $\partial u/\partial y\,(c,0)$ is bounded.

solution of Laplace's equation on both sides of the boundary. Thus it generalises the result that, if

$$\alpha(t)w_+(t) + \beta(t)w_-(t) = 0$$

on a closed boundary Γ, then the total number of zeros of $w_+(z)$ and $w_-(z)$ in their domains of definition is the change in the argument of $-\alpha/\beta$ as we go around Γ. Clearly, the change of argument provides a global measure of the 'strengths', i.e. the powers of z that control the local behaviour at all the singularities on the boundary for a solution of Laplace's equation, as described in detail in [20]. Riemann–Hilbert theory can be used to reduce the problem to that of a singular integral equation, and then index theory can be thought of as a generalisation of the Fredholm Alternative as applied to conventional integral equations with square-integrable kernels [42]: the statement is simply that there is only a unique solution when the sum of the powers of r (the distance from any particular singularity) occurring in the local behaviour of the solution is equal to the index.

*5.10 Localised boundary data

Because singularities in the solutions of elliptic boundary value problems do not propagate as they do for hyperbolic problems, there can be no counterpart in this chapter of the analysis of the response to localised Cauchy data that we gave in §4.5.5 and Exercise 4.22. However, we can pose the following interesting and practically useful question:

> 'Suppose the solution of an elliptic boundary value problem varies by $O(1)$ or less in a region D of dimension $O(1)$ or greater. Now introduce Dirichlet or Neumann conditions on the boundary ∂D_ϵ of a small subset of D. Does this cause the solution to change by $O(1)$ except in a region close to ∂D_ϵ?'

For example, suppose D is the unit disc and the harmonic function $u(x, y)$ vanishes on its boundary. Then $u \equiv 0$. However, if we additionally require $u = 1$ on $x^2 + y^2 = \epsilon^2$, then the solution is $u = \log(x^2 + y^2)/(2 \log \epsilon)$, which is small unless $x^2 + y^2$ is small. We would need to impose Dirichlet data of $O(\log \epsilon)$, or Neumann data of $O(1/\epsilon)$, in order to produce an $O(1)$ change in u away from ∂D_ϵ. However, our methodology for models of cracks and aerofoils tells us that, if D is $\mathbb{R}^2$ and $u = 1$ on the curves $y = \epsilon f_\pm(x)$ near $y = 0$, $0 < x < 1$, with u bounded on ∂D_ϵ and $u = O(1)$ at infinity, then, as $\epsilon \to 0$, $u \to (\theta_1 + \theta_2)/\pi$ in terms of polar angles θ_1 and θ_2 centred at $(0, 0)$ and $(1, 0)$, respectively. Hence, even though the region excluded from D has zero area as $\epsilon \to 0$, the imposition of $O(1)$ Dirichlet data on its boundary does now change the solution by $O(1)$ even when we are not close to the boundary.

The jargon appropriate to this situation comes from the electrostatic model of §5.1. For the case of constant Dirichlet data u_0, the ratio

$$\frac{1}{u_0} \int_{\partial D_\epsilon} \frac{\partial u}{\partial n} \, ds$$

is called the *capacity* of ∂D_ϵ; as $\epsilon \to 0$ it is $O(|\log \epsilon|^{-1})$ in the first example and $O(1)$ in the second. It can be shown more generally that, for Laplace's equation, a point in two dimensions has zero capacity and a smooth curve has finite capacity. Equally, in three dimensions a point or a smooth curve has zero capacity but a smooth surface has finite capacity. The long-range influence of more convoluted boundaries ∂D_ϵ is an important question that often arises in studies of the effects of surface roughness. This is best answered using the asymptotic method of homogenisation, and the theory of capacity is described in detail in [33].

5.11 Nonlinear problems

Most of the problems considered in this chapter so far have been linear and most of the methods we have discussed only apply to linear problems. Despite the heterogeneity of the results, compared say to the unity of Chapter 4, the general picture that emerges is that those linear problems that we have been able to analyse are either well posed (unless, of course, we were solving with Cauchy data), or fail to have solutions except in special cases. This is an inevitable consequence of the Fredholm Alternative, even though the presence of partial derivatives makes matters much more complicated technically than for linear algebraic equations or linear ordinary differential equations. In particular, the presence of singularities in the data, either on the boundary or in the coefficients of the differential equations, calls for great care.

We now wish to work towards a scenario of what might be expected to happen when nonlinearity is introduced, and in particular to study the influence of the nonlinear terms as they get larger and larger in comparison with the linear ones. It is very likely that problems that are well posed linearly are still well posed for small enough nonlinearity, but much more dramatic behaviour can be expected in other cases. However, we are even less likely to be able to rely on explicit solutions to nonlinear equations than we are for linear ones. Hence most of this section deals with methods that lead to general existence, uniqueness and smoothness results. We begin by briefly recalling a few of the common practical situations that can only be modelled by nonlinear elliptic equations.

5.11.1 Nonlinear models

An exothermic chemical reaction produces heat at a rate which depends on the temperature T, normally via an 'Arrhenius' function $f(T) \propto e^{-E/RT}$, where E and R are constants. If the heat flux is equal to $-\nabla T$, then T satisfies

$$\nabla^2 T + f(T) = 0. \tag{5.152}$$

Whether or not such a steady-state model can have solutions depends physically on whether heat conduction can remove the heat produced by the reaction quickly enough. A similar model applies to reactions controlled by concentration, with T replaced by the reaction concentration c. When $f(c) > 0$ such a reaction is called *autocatalytic*.

In fluid mechanics, when we consider steady two-dimensional inviscid flow, without the assumption of irrotationality that we made in §5.1.4.1, we find that if we define the vorticity

$$\boldsymbol{\omega} = \nabla \wedge (u(x,y), v(x,y), 0) = (0, 0, \omega(x,y))$$

then the curl of (2.6), with ρ taken constant, becomes $u\,\partial\omega/\partial x + v\,\partial\omega/\partial y = 0$. Hence, since the stream function ψ is such that $u\,\partial\psi/\partial x + v\,\partial\psi/\partial y = 0$ and $\omega = -\nabla^2\psi$,

$$\nabla^2\psi + f(\psi) = 0,$$

where f is some function to be determined by the problem, rather than being prescribed directly as in (5.152).[92]

As in §5.1.3, it is not hard to imagine generalisations to include convective terms which introduce first-order derivatives into (5.152), or to vector-valued dependent variables.

In a different vein, we have already remarked in §4.8 that subsonic steady inviscid gas flow leads to a quasilinear elliptic equation for the velocity potential which can be linearised by the hodograph method. Also, very simple generalisations of some of the situations listed in §5.1 lead to other quasilinear examples. For example, suppose the Darcy flow in §5.1.4.4 had been of a compressible gas rather than an incompressible liquid. Then the mass conservation equation $\nabla\cdot(\rho\mathbf{v}) = 0$ gives

$$\nabla\cdot(\rho(P)\nabla P) = 0 \tag{5.153}$$

for an isothermal flow in which ρ is a prescribed function of the pressure P. Equally, unidirectional flow of a variable-viscosity incompressible fluid with velocity $(0, 0, w(x, y))$ only satisfies the slow flow equations (5.24) and (5.25) if[93]

$$\nabla\cdot(\mu\nabla w) = 0; \tag{5.154}$$

when the fluid is non-Newtonian we can sometimes set $\mu = \mu(\nabla w)$. Note that (5.153) can be transformed into Laplace's equation by using the *Kirchhoff transform* $u = \int^P \rho(P')\,\mathrm{d}P'$. However, (5.154) is not so easily linearisable; when $\mu = |\nabla w|^{p-2}$ it is called the *p-Laplace equation*, and the commonly occurring Darcy flow in which the velocity–pressure law takes the nonlinear form $|\mathbf{v}|\mathbf{v} = -k\nabla P$ leads to the 3/2-Laplace equation for P. An elementary calculation (see Exercise 5.43) shows that the p-Laplace equation is only elliptic if $p > 1$, so one must not be deceived into thinking that all conservation laws of the form (5.154) are automatically elliptic.[94]

5.11.2 Existence and uniqueness

5.11.2.1 Comparison methods

For semilinear equations we can use consequences of the maximum principle as in §5.3 to find solutions as limits of sequences of approximations from above or below. Consider, for simplicity, an equation of the type

[92]The same model arises in connection with plasma confinement [17], and it will be referred to again in Chapter 7.

[93]This is also the condition for the Navier–Stokes equations to be satisfied, as we will see in Chapter 9.

[94]Note that the current flow in an electric window heater could be modelled by $\nabla\cdot(\sigma\nabla\phi) = 0$, where ϕ is the potential and σ is the conductivity (the reciprocal of the resistivity); if we wish to have uniform electric heating, then the product of the current density $\sigma\nabla\phi$ and the electric field $-\nabla\phi$ must be constant, which leads to the *hyperbolic* equation $\nabla\cdot(\nabla\phi/|\nabla\phi|^2) = 0$.

$$\nabla^2 u + f(u) = 0 \quad \text{in } D, \tag{5.155}$$

with Dirichlet data

$$u = g(x) \quad \text{on } \partial D, \tag{5.156}$$

although the method works equally well with Robin boundary conditions. The function f is assumed to be Lipschitz continuous in u. Motivated by the comparison theorem in §5.3, a function $\overline{u}(x)$ is said to be an ***upper solution*** (or ***supersolution***) if it satisfies

$$\nabla^2 \overline{u} + f(\overline{u}) \leqslant 0 \quad \text{in } D \qquad \text{and} \qquad \overline{u} \geqslant g(x) \quad \text{on } \partial D.$$

Equally, if $\underline{u}$ satisfies these with the inequalities reversed, it is called a ***lower solution*** (or ***subsolution***). If we can find such a pair $\underline{u}$ and $\overline{u}$, and if the important inequality $\underline{u} \leqslant \overline{u}$ is satisfied, then there is sometimes a constructive proof that there is at least one solution of (5.155) and (5.156) lying between them. Suppose, for example, that f in (5.155) is such that there is a $K \geqslant 0$ for which $F(u) \equiv f(u) + Ku$ is an increasing function for all u of interest. Since

$$\nabla^2 u - Ku + F(u) = 0,$$

we proceed iteratively, defining $u_0 = \underline{u}$ and

$$\nabla^2 u_n - Ku_n + F(u_{n-1}) = 0 \quad \text{in } D,$$

with

$$u_n = g(x) \quad \text{on } \partial D$$

for $n \geqslant 1$. Now, by generalising the maximum principle to cover the differential inequality $\nabla^2 u - Ku < 0$, and using the properties of F, it can be shown that the functions u_n form an increasing sequence of lower solutions, bounded above by $\overline{u}$. These necessarily converge to some u, which is a solution of the original problem.

Another important use of upper and lower solutions is in estimating the size of solutions.

Example 5.1 Suppose that $\nabla^2 u + f(u) = 0$ in D, $u \geqslant 0$ on ∂D, $f(u) \geqslant 1$ and the sphere $|\mathbf{x}| \leqslant 1$ is inside D. Can we find a lower bound for u in D?

We only have to look at the simple problem

$$\nabla^2 v + 1 = 0 \quad \text{in } |\mathbf{x}| < 1 \qquad \text{with} \qquad v = 0 \quad \text{on } |\mathbf{x}| = 1. \tag{5.157}$$

Clearly, $v = (1 - |\mathbf{x}|^2)/2m$, where m is the spatial dimension, and u is an upper solution for (5.157). Thus any solution u is never less than $(1 - |\mathbf{x}|^2)/2m$.

5.11.2.2 Variational methods

Several of the remarks make in §5.4 carry over to nonlinear elliptic problems in the event that they are Euler–Lagrange equations. Such is the case, say, for the p-Laplace equation, whose solutions give stationary points of the functional

$$J(u) = \int_D |\nabla u|^p \, d\mathbf{x}. \tag{5.158}$$

One especially interesting and well-studied variational problem concerns the calculation of ***minimal surfaces***, namely surfaces $u = u(x, y)$ such that the area

$$\iint_D \left(1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right)^{1/2} \mathrm{d}x\,\mathrm{d}y$$

is minimised. The Euler–Lagrange equation,

$$\left(1 + \left(\frac{\partial u}{\partial y}\right)^2\right)\frac{\partial^2 u}{\partial x^2} - 2\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}\frac{\partial^2 u}{\partial x\,\partial y} + \left(1 + \left(\frac{\partial u}{\partial x}\right)^2\right)\frac{\partial^2 u}{\partial y^2} = 0,$$

is elliptic; for small u, it is approximated by Laplace's equation for a membrane, as described in §5.1.4.2.

Theoretically, the most important attribute of weak or variational methods for elliptic problems is that they may allow us to study situations in which singularities occur, and we will discuss this further in the next section. The philosophy is exactly as for the theory of weak solutions of hyperbolic equations in Chapter 2, although, of course, the form of the singularities is very different in the two cases. We will shortly encounter several such situations, but we will eschew the functional analysis that is needed to enable variational methods to be brought to bear on what are often very delicate problems. For example, it can easily happen that, for the generalisation of (5.158) to $\int_D (|\nabla u|^p - F(u))\,\mathrm{d}x$, F is so large that $J(u)$ only has $-\infty$ as a lower bound.

5.11.3 Parameter dependence and singular behaviour

Comparison methods can sometimes be used to prove that solutions depend continuously on the data and we expect this to be the case for most elliptic equations with appropriate boundary conditions. However, when large changes in the data occur, there can be surprising behaviour, which has far-reaching implications for many practical problems.

5.11.3.1 Nonlinear eigenvalue problems

Nonlinear elliptic boundary value problems are frequently posed as ***nonlinear eigenvalue problems***. For instance, an approximation to the exothermic chemical reaction problem gives rise to the equation[95]

$$\nabla^2 u + \lambda \mathrm{e}^u = 0 \quad \text{in } D. \tag{5.159}$$

The parameter λ can be thought of as a non-dimensional heat of reaction or reactant concentration. In a typical situation of interest with Dirichlet conditions, say

$$u = 0 \quad \text{on } \partial D, \tag{5.160}$$

we might want to know whether or not there is a solution and, if there is, whether it is unique.

[95] When $\lambda < 0$, this equation is the elliptic form of Liouville's equation; it also appears in differential geometry for the 'kernel function' and we will see it again in §5.12 and Chapter 6.

Taking λ to be sufficiently small, corresponding to a low rate of reaction, we can see that the problem (5.159) and (5.160) has a solution in which u is everywhere small. One way to do this is to use upper and lower solutions. Clearly, $\underline{u} \equiv 0$ satisfies the boundary condition, while $\nabla^2 \underline{u} + \lambda e^{\underline{u}} > 0$, and so zero is a lower solution. Now we write $\overline{u} = \mu w$, where μ is a positive constant and where w satisfies a trivial Poisson equation:

$$\nabla^2 w + 1 = 0 \quad \text{in } D \qquad \text{with} \qquad w = 0 \quad \text{on } \partial D.$$

We see that $\overline{u}$ satisfies the boundary condition and $\nabla^2 \overline{u} + \lambda e^{\overline{u}} = \lambda e^{\mu w} - \mu \leqslant 0$, provided that $\mu e^{-\mu w} \geqslant \lambda$ in D. Hence $\overline{u}$, which is positive since $w \geqslant 0$, is an upper solution if

$$\lambda \leqslant \mu \exp\left(-\mu \sup_D\{w\}\right).$$

It follows that for $\lambda \leqslant (e \sup_D\{w\})^{-1}$ there is a positive solution.

An alternative existence proof follows from the *contraction mapping theorem.* This relies upon the existence of a mapping, $\mathcal{T}$, taking one function defined on D to another, such that, in a suitable norm, $\|\mathcal{T}(v-w)\| < k\|v-w\|$ for some $k < 1$. The fixed point of $\mathcal{T}$ gives the solution. In this case, if $\mathcal{T}$ is defined to map w into v, where

$$\nabla^2 v + \lambda e^{w} = 0 \qquad \text{with} \qquad v = 0 \quad \text{on } \partial D,$$

then the proof based on the iteration $u_n = \mathcal{T} u_{n-1}$ is seen to be essentially the same as that using upper and lower solutions.

The Helmholtz problem (5.27) can be used to demonstrate non-existence of solutions to (5.159) and (5.160). We know from §5.7.1 that we can take ϕ and μ to be the positive principal eigenfunction and eigenvalue of $-\nabla^2$, respectively, so that

$$-\nabla^2 \phi = \mu \phi \quad \text{in } D \qquad \text{with} \qquad \phi = 0 \quad \text{on } \partial D;$$

if we define the Fourier coefficient $a = \int_D \phi u \, dx$, integration by parts shows that

$$a = \frac{\lambda}{\mu} \int_D \varphi e^{u} \, dx \geqslant \frac{\lambda}{\mu} I e^{a/I},$$

where $I = \int_D \varphi \, dx$; in the last step we have used Jensen's inequality.[96] For $\lambda > \mu/e$, no a can satisfy the inequality and no solution can exist for such values of λ. This, and similar, nonlinear eigenvalue problems have a bounded spectrum, by which we mean there is some λ^*, which is less than μ/e in the above example, such that there is at least one solution for $\lambda < \lambda^*$ but no solution for $\lambda > \lambda^*$.

[96] Jensen's inequality states that, if $w(x) \geqslant 0$ is such that $\int_D w \, dx = 1$, then

$$\int_D w f(u) \, dx \geqslant f\left(\int_D w u \, dx\right)$$

for arbitrary smooth u, provided that f is convex, i.e.

$$f(\alpha a + (1-\alpha) b) \leqslant \alpha f(a) + (1-\alpha) f(b)$$

for $0 \leqslant \alpha \leqslant 1$ (see Exercise 5.51).

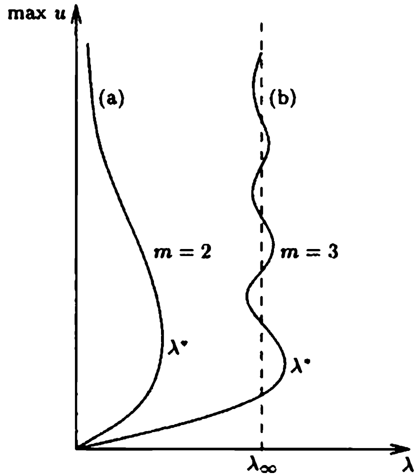


Fig. 5.7 Response diagrams for (5.161).

To see more precisely how solutions depend upon λ, it is easiest to start with the special case in which D is the unit ball in m dimensions. If we assume u is radially symmetric so that $u = u(r)$ in polar coordinates,[97] the problem reduces to the ordinary differential equation

$$\frac{\mathrm{d}^2u}{\mathrm{d}r^2} + \frac{m-1}{r}\frac{\mathrm{d}u}{\mathrm{d}r} + \lambda \mathrm{e}^u = 0, \tag{5.161}$$

with $u(1) = 0$ and a regularity condition at $r = 0$, namely that $\mathrm{d}u/\mathrm{d}r\,(0) = 0$ for $m = 1$ and u is bounded for $m > 1$. On making the change of variables $r = \mathrm{e}^s$ and $u = v - 2\log s$ (which is only needed for $m > 1$), the differential equation becomes autonomous. More details of this transformation are given in Exercise 5.47, but the fact that it allows us to reduce the problem to one in a phase-plane allows us to plot the *response diagrams* of, say, the maximum value of u as a function of λ, and these diagrams are found to be as in Fig. 5.7. Amongst the interesting features are non-uniqueness and unboundedness; when we identify the solution as an equilibrium state for a parabolic problem in Chapter 6, we will also find a change of stability at $\lambda = \lambda^*$. From an applied mathematical viewpoint, stability usually needs to be discussed in the framework of an evolutionary model, which is precisely what we will do in §6.6.4. However, for elliptic problems that are Euler–Lagrange equations it is possible to conjecture stability results such as these without appeal to time-dependent generalisations. In such cases, if it can be shown that a solution branch is a global minimiser of the energy, the solution is likely to be stable as a steady state of any reasonable evolution model. If we are lucky, the link between

[97]That this has to be the form of u can be proved rigorously in certain circumstances (see Exercise 5.46).

the 'time-dependent' and 'energy' concepts can be made by using the energy as a Lyapunov function, as in the theory of ordinary differential equations.

This discussion provokes two other questions. The first concerns the possibility of 'spontaneous' singularities appearing in the solution domain.

5.11.3.2 Singularities

We know from our general discussion in Chapter 3 that singularities in the solution of elliptic problems cannot propagate, in that they cannot exist on manifolds of dimension one fewer than the space of independent variables, but there is nothing to prevent more isolated singularities occurring. Indeed, for $m \geqslant 3$ in (5.159) and (5.160), it can be shown that there is a value λ_∞ such that there are solutions u and parameters λ with u tending to infinity at some interior point of D as $\lambda \to \lambda_\infty$. Additionally, the singular solutions U are weak in the sense that

$$\int_D \nabla\psi \cdot \nabla U \, \mathrm{d}\mathbf{x} = \lambda_\infty \int_D \psi \mathrm{e}^U \, \mathrm{d}\mathbf{x}$$

for all test functions ψ in some suitable space. We can see this explicitly for the symmetric case[98] (5.161), when $\lambda_\infty = 2(m-2)$ and $U = -2\log r$.

The discussion above raises the question of the relationship between the possible forms of the response diagram when $u_{\max}$ is large and the growth of f for equations of the form

$$\nabla^2 u + \lambda f(u) = 0, \tag{5.162}$$

with, say, zero Dirichlet data. In particular, if f grows as a power as $u \to \infty$, how does the response depend on this power and the dimension m? It is easily seen by direct integration that, when $m = 1$ and $f(u)/u \to \infty$ as $u \to \infty$, then there are large solutions for small λ, as in Fig. 5.7(a). More generally, and motivated by the analysis of (5.161), it can be shown that, for smooth, bounded, m-dimensional regions D, the following scenarios are possible.

1. If f grows linearly, with $f(u)/u \to K > 0$ as $u \to \infty$, there are large solutions for λ close to μ/K, where μ is the principal eigenvalue for

$$\nabla^2\phi + \mu\phi = 0 \quad \text{in } D \qquad \text{with} \qquad \phi = 0 \quad \text{on } \partial D. \tag{5.163}$$

2. If the growth is sublinear, so that $f(u)/u \to 0$ as $u \to \infty$, and $f(u) > 0$, then u exists and $u_{\max}$ is large for large values of λ.
3. If $f(u)$ is identically a power, say $f(u) = u^p$ with $p > 1$, the form of the response diagram depends crucially upon the relation between p and m. In this case, there is always the trivial solution $u \equiv 0$. For $m = 1$ or 2, or $p < (m+2)/(m-2)$ with $m \geqslant 3$, there is a non-trivial solution for all λ. For $p \geqslant (m+2)/(m-2)$ with $m \geqslant 3$, there is no non-trivial solution. The special value, $p = p_c \equiv (m+2)/(m-2)$, is called the *critical Sobolev exponent*. This case is distinguished in that it allows

[98] It can be shown that $\lambda_\infty < \lambda^*$ for $3 \leqslant m \leqslant 9$, and $\lambda_\infty = \lambda^*$ for $m \geqslant 10$. In two dimensions there is an apparently similar situation, where there is a solution satisfying the equation for $0 < r < 1$ and the condition at $r = 1$, and having logarithmic growth as $r \to 0$. This singular solution is *not* a weak solution.

us to proceed more easily in solving the equation that corresponds to (5.161) via a phase-plane analysis (see Exercise 5.48).

5.11.3.3 Non-uniqueness and bifurcations

The discussion above has revealed how easy it is for solution branches of nonlinear elliptic equations to cease to exist or tend to infinity at critical values of the control parameter λ. Another kind of pathology is that of *bifurcation*, by which we mean the branching of a new solution from another 'reference' solution as the parameter is varied.

These kinds of behaviour can be approached systematically by careful examination of the local dependence of the solution on the parameter. Let us begin with the simplest case of the nonlinear eigenvalue problem

$$\nabla^2 u + \lambda f(u) = 0 \quad \text{in } D \qquad \text{with} \qquad u = 0 \quad \text{on } \partial D, \tag{5.164}$$

in which $f(0) = 0$ and $f'(0) \neq 0$, and let us see whether any solutions can exist near the trivial solution $u = 0$.[99] The key step is to search in the vicinity of a particular value of λ, say λ_0, by writing $\lambda - \lambda_0 = \epsilon$ and, in the first instance, expanding u in the form

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \cdots ; \tag{5.165}$$

thus we are effectively seeking the derivative of u with respect to λ. Equating the coefficients of like powers of ϵ, we soon see that

$$\nabla^2 u_1 + \lambda_0 f'(0) u_1 = 0 \quad \text{in } D \qquad \text{with} \qquad u_1 = 0 \quad \text{on } \partial D,$$

so that u_1 can only be non-zero if $-\lambda_0 f'(0)$ is an eigenvalue of the Laplacian in D, with corresponding normed eigenfunction ϕ. Writing $u_1 = \alpha\phi$, where $\alpha \neq 0$, we find that u_2 satisfies

$$\nabla^2 u_2 + \lambda_0 f'(0) u_2 + \frac{1}{2}\lambda_0 \alpha^2 f''(0)\phi^2 + \alpha f'(0)\phi = 0. \tag{5.166}$$

Hence, by the Fredholm Alternative, u_2 can only exist if α satisfies

$$\frac{1}{2}\alpha\lambda_0 f''(0) \int_D \phi^3 \, \mathrm{dx} + f'(0) \int_D \phi^2 \, \mathrm{dx} = 0. \tag{5.167}$$

If $\int_D \phi^3 \, \mathrm{dx} \neq 0$, we then have what is called a ***transcritical bifurcation*** from the zero solution. The bifurcation solution is locally, for λ near the eigenvalue λ_0, an eigenfunction $\alpha\phi$ of known amplitude. Moreover, the discussion of §5.7.1 shows that bifurcation at the principal eigenvalue leads to one-signed bifurcation solutions (see Exercise 5.49).

[99] When $f(u) = \sin u$, (5.164) is a model for the buckling of an elastic strut (the 'Euler strut'), where u is the transverse displacement and λ is the compressive load applied along the strut.

Now let us look at cases of (5.164) where we have a non-trivial reference solution $u_0(\mathbf{x}, \lambda)$. Adopting the same notation as in (5.165), we now find

$$\nabla^2 u_1 + \lambda_0 f'(u_0(\mathbf{x}))\, u_1 = -f(u_0(\mathbf{x})) \quad \text{in } D \qquad \text{with} \qquad u_1 = 0 \quad \text{on } \partial D. \tag{5.168}$$

Hence there is a unique solution for u_0 unless λ_0 is an eigenvalue of the (less trivial, because of the $\mathbf{x}$ dependence in the coefficient of u_1) eigenvalue problem in which the right-hand side of (5.168) is set equal to zero. If λ_0 is such an eigenvalue, there are two possibilities.

The most likely is that $f(u_0)$ is not orthogonal to the eigenfunction ϕ corresponding to λ_0. In this case u_1 does not exist and any solutions that are 'close' to u_0 cannot be found by the prescription (5.165). What this means is that we must seek a more general representation for u, say in the form

$$u = u_0 + |\epsilon|^{1/2} u_1 + \epsilon u_2 + \cdots, \tag{5.169}$$

and this is precisely the behaviour we saw near the turnover points $\lambda = \lambda^*$ in the previous section. Locally the solution depends on $|\lambda - \lambda_0|^{1/2}$, and no bifurcation occurs. Indeed, we could say that $\lambda - \lambda_0$ is a locally smooth function of $\max_D u_0$.

The second possibility is that

$$\int_D \phi(\mathbf{x}) f(u_0(\mathbf{x}))\, d\mathbf{x} = 0.$$

In this case, by the Fredholm Alternative, there is a continuum of solutions of (5.168). Noting that the derivative of (5.164) with respect to λ gives that

$$\nabla^2 \frac{\partial u_0}{\partial \lambda} + \lambda f'(u_0) \frac{\partial u_0}{\partial \lambda} = -f(u_0),$$

we see that $\partial u_0/\partial\lambda$ is a particular integral of (5.168) which vanishes on ∂D, and hence that

$$u_1 = \frac{\partial u_0}{\partial \lambda} + \alpha\phi,$$

where α is again an arbitrary constant. The problem for u_2 is now

$$\nabla^2 u_2 + \lambda_0 f'(u_0) u_2 + \left(\frac{\partial u_0}{\partial \lambda} + \alpha\phi\right)\left(f'(u_0) + \frac{1}{2}\lambda_0 f''(u_0)\left(\frac{\partial u_0}{\partial \lambda} + \alpha\phi\right)\right) = 0.$$

However, by differentiating (5.164) twice with respect to λ and using the Fredholm Alternative in reverse, we find that $(\partial u_0/\partial\lambda) f'(u_0) + (\lambda_0/2)(\partial u_0/\partial\lambda)^2 f''(u_0)$ is orthogonal to ϕ, and hence a third application of the Alternative to (5.166) shows that either $\alpha = 0$ or

$$\alpha = -\int_D \phi^2 \left(f'(u_0) + \lambda_0 f''(u_0)\frac{\partial u_0}{\partial \lambda}\right) d\mathbf{x} \bigg/ \int_D \frac{1}{2}\lambda_0 \phi^3 f''(u_0)\, d\mathbf{x}. \tag{5.170}$$

The case $\alpha = 0$ just means that (5.165) is the Taylor expansion of $u_0(\mathbf{x}, \lambda)$ about $\lambda = \lambda_0$, and is analogous to the smooth continuation of the zero solution in the case $f(0) = 0$. Hence (5.170) corresponds to another transcritical bifurcation.

We notice that either (5.167) or (5.170) fails when certain integrals vanish ($\int_D \phi^3 f''(u_0)\,\mathrm{dx}$ for (5.170)), and in this case we again have to revert to a representation like (5.169). However, this will not now lead to a turnover in the response curve, but rather a 'transverse' bifurcation in which the bifurcating solution branch is perpendicular to the reference branch; this is called a *pitchfork bifurcation*.

We conclude with the observation that sometimes it is possible to show that at most one solution exists to (5.164). We illustrate this by showing that there is at most one positive solution if $f(u)/u$ is a strictly decreasing function. We must make the assumption that all solutions are smooth, and, if there are two solutions, then they intersect at a reasonably regular surface. We suppose that u and v are two distinct, positive solutions. Then $u - v$ is positive in some region D_+ and vanishes on ∂D_+. Integration by parts over D_+ yields

$$\int_{D_+} (vf(u) - uf(v))\,\mathrm{dx} = \int_{\partial D_+} u\left(\frac{\partial v}{\partial n} - \frac{\partial u}{\partial n}\right) ds.$$

Since the left-hand side is negative while the right-hand side is non-negative, the assumption is contradicted.

5.11.3.4 Other irregular behaviour

We have already noted, in §5.11.3.2, that solutions to nonlinear equations may be infinite at some points whose position is not known in advance, unlike solutions of linear problems whose interior singularities are determined by the coefficients in the equations. However, two weaker types of irregularity can arise. For example, if we look at the radially-symmetric solution of $\nabla \cdot (|\nabla u|\nabla u) - 1 = 0$, then $|\nabla u| = O(r^{3/2})$ for $r \to 0$. The solution, although differentiable everywhere, is not twice differentiable at the origin. Another type of behaviour, which will be considered further in Chapter 7, is exhibited by semilinear problems in which $f(u)$ is not Lipschitz continuous, such as

$$\nabla^2 u = u^p \quad \text{in } D \qquad \text{with} \qquad u = g > 0 \quad \text{on } \partial D$$

and $0 < p < 1$. If the region D is big enough, it can be shown (see Exercise 5.50) that there is a 'dead core', defined to be a region contained in D, say D_0, in which $u \equiv 0$. Outside D_0, u is positive. On the boundary ∂D_0, $\partial u/\partial n$ as well as u is zero, and clearly this solution is not analytic on ∂D_0.

5.12 Liouville's equation again

We conclude this chapter by pointing out a remarkable relationship between the elliptic version of Liouville's equation in two dimensions and Green's functions. Suppose we return to §5.9.1, and consider the problem for the Green's function $G(z, z_0)$ for Laplace's equation in a closed region with Dirichlet boundary data. Let us assume that the region can be mapped conformally onto $|\zeta| \leqslant 1$ by the map

$$\zeta = f(z).$$

We assumed earlier that the point z_0 is mapped onto $\zeta = 0$, so the mapping is different for each choice of z_0. Here, we will let z_0 vary, so we assume that this point is mapped to ζ_0 and keep f fixed. Then, using the mapping of (5.117) to map the unit disc onto itself, the Green's function is simply

$$G(z,z_0) = \frac{1}{2\pi}\log\left|\frac{\zeta-\zeta_0}{1-\zeta\bar{\zeta}_0}\right|$$
$$= \frac{1}{2\pi}\left(\log|z-z_0| + H(z,z_0)\right),$$

say. Now set $H(z,z) = T(z)$ so that,[100] taking the limit $z \to z_0$,

$$T(z) = \log\left(\frac{|f'(z)|}{1-|f(z)|^2}\right). \tag{5.171}$$

Then it is easy to show that

$$\nabla^2 T(z) = -\nabla^2 \log(1-|f(z)|^2)$$
$$= -4|f'(z)|^2 \frac{\partial^2}{\partial\zeta\,\partial\bar{\zeta}} \log(1-\zeta\bar{\zeta})$$
$$= 4e^{2T},$$

which, with a trivial change of variable, is Liouville's equation (5.159). Clearly, the argument above suggests that (5.171) is the general solution to this nonlinear equation, but we knew this anyway from the discussion in §4.8.3. Writing $u = 2T + \log 2$ and

$$\nabla^2 u = 4e^u$$

as

$$\frac{\partial^2 u}{\partial z \partial \bar{z}} = e^u,$$

and formally setting $X = if(z)$ and $Y = i\overline{f(z)}$ and $\gamma = i\sqrt{2}$, we find that

$$u = 2\log\left(\frac{2|f'(z)|}{1-|f(z)|^2}\right),$$

which is (5.171)!

5.13 Postscript: ∇^2 or $-\Delta$?

Many textbooks, especially the more theoretical ones, use the symbol Δ for the Laplace operator, and write Laplace's equation as

$$-\Delta u = 0$$

rather than $\nabla^2 u = 0$. Apart from the different notations for the Laplacian, there are theoretical reasons for introducing the minus sign, as it automatically ensures

[100]The function $H(z,z_0)$, which can be loosely called the 'regular part of the Green's function', has many important applications (see Exercise 5.52). The function $T(z)$ is known as the Bergman kernel function.

positivity of many quantities of central importance. For example, when Poisson's equation is written $-\Delta u = f$, with $u = 0$ on ∂D, positive data f gives a positive solution u. Likewise, the Green's function for the operator $-\Delta$ with homogeneous boundary data is positive, as are its eigenvalues and 'Fourier symbol' $|\mathbf{k}|^2$. It is also often natural to think of an elliptic equation as arising out of the large-time behaviour of an evolution problem such as the parabolic equation $\partial u/\partial t - \nabla^2 u = f$, which again leads to $-\Delta u = f$. Set against this, the vast majority of 'end-users' of partial differential equations conventionally write ∇^2 and think of Green's and Riemann functions as the solutions of '$\mathcal{L}u = \delta$', and we have followed that convention in this book.

Exercises

5.1. Show that (5.23) and (5.24) are compatible with the existence of an *Airy stress function* $A(x, y)$ such that

$$\frac{\partial^2 A}{\partial y^2} = \lambda\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + 2\mu\frac{\partial u}{\partial x} = \sigma_x,$$

$$\frac{\partial^2 A}{\partial x^2} = \lambda\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + 2\mu\frac{\partial v}{\partial y} = \sigma_y,$$

$$-\frac{\partial^2 A}{\partial x \partial y} = \mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) = \tau,$$

and that $\nabla^4 A = 0$. The condition that the traction on a boundary $y = f(x)$ is zero can be shown to be

$$\sigma_x \frac{\mathrm{d}f}{\mathrm{d}x} - \tau = \tau\frac{\mathrm{d}f}{\mathrm{d}x} - \sigma_y = 0.$$

Show that these conditions imply that

$$A = \text{constant}, \quad \frac{\partial A}{\partial n} = 0 \quad \text{on the boundary.}$$

(In a simply-connected domain the constant can be taken to be zero without loss of generality.)

5.2. Show that the boundary value problem

$$\nabla^2 u = c = \text{constant} \quad \text{for } x^2 + y^2 = r^2 < 1,$$

with

$$\frac{\partial u}{\partial r} = 2 \quad \text{on } r = 1,$$

only has solutions for $c = 4$ and that, in this case, $u = r^2 + a$ is a solution for an arbitrary constant a.

5.3. Suppose that $\nabla^2 u = 0$, $1 < r < 2$, in two-dimensional polar coordinates, with

$$\frac{\partial u}{\partial r} + \alpha_1 u = k\cos\theta \quad \text{on } r = 1 \qquad \text{and} \qquad \frac{\partial u}{\partial r} + \alpha_2 u = 0 \quad \text{on } r = 2.$$

Seek a solution in which u is a function of r multiplied by $\cos\theta$, and show that it exists and is the only one of this form unless $6\alpha_1\alpha_2 + 5\alpha_1 - 10\alpha_2 - 3 = 0$. Repeat the calculation when $\cos\theta$ is replaced by $\cos n\theta$. Deduce that there is a unique solution to any of these problems unless α_1 and α_2 satisfy a countably infinite number of conditions.

5.4. Show that, if

$$\nabla^2 u = 0 \quad \text{for } -\pi < x < \pi,\ -\pi < y < \pi,$$

with

$$u = 0 \quad \text{on } y = 0, \pi \qquad \text{and} \qquad \frac{\partial u}{\partial x} = \pm\gamma u \quad \text{on } x = \pm\pi, \text{ respectively},$$

then there is a non-trivial solution when $\gamma \tanh n\pi = n$, where n is an integer.

5.5. Suppose $\nabla^2 u = 0$ in a square and that $u = 1$ on one side of the square, $u = 0$ on the other three sides, and u is bounded. Show that $u = \frac{1}{4}$ at the centre of the square. What is the corresponding result for a cube?

5.6. Suppose

$$\nabla^2 u - cu = f \quad \text{in } D,$$

with

$$u = g \quad \text{on } \partial D,$$

where D is bounded and ∂D is smooth. If u exists, show that it is unique if $c > 0$. If $c < 0$ and $f = g = 0$, and D is the region $r^2 = x^2 + y^2 < 1$, show that

$$u = \text{constant} \cdot J_0\left(r\sqrt{-c}\right)$$

is a solution as long as $\sqrt{-c}$ is a zero of the Bessel function $J_0(x)$, which satisfies

$$\frac{d^2 J_0}{dx^2} + \frac{1}{x}\frac{dJ_0}{dx} + J_0 = 0.$$

If $c < 0$ and $f = g = 0$, and D is the region $1 < x^2 + y^2 + z^2 < 4$, show that there are non-trivial solutions if $c = -n^2\pi^2$, $n = 1, 2, 3, \ldots$.

5.7. By setting $\boldsymbol{\xi} = \mathbf{0}$ in (5.59), show that a solution of Laplace's equation in two dimensions satisfies

$$u(\mathbf{0}) = \frac{1}{2\pi a}\int_{\partial D} u \, ds,$$

where ∂D is a circle with centre $\mathbf{0}$ and radius a. This is called the *mean value theorem*. Derive the maximum principle from this result. What is the corresponding result in three dimensions?

5.8. Why does the maximum principle not hold for

$$\mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + \lambda^2 \quad \text{for } 0 \leqslant x \leqslant 1,$$

with $\lambda > \pi$? Show a similar result for $\mathcal{L} = \nabla^2 + \lambda^2$, $0 \leqslant x^2 + y^2 + z^2 \leqslant 1$, by using the fact that $\sin(\pi r)/r$ satisfies $\mathcal{L}u = 0$, with $u = 0$ on $r = 1$.

5.9. Suppose that $u(x, y)$ tends to zero as rapidly as you wish as $r^2 = x^2+y^2 \to \infty$ and $\nabla^2 u$ is integrable everywhere. Show that

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u\nabla^2 \log(r^2+\epsilon)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \log(r^2+\epsilon)\nabla^2 u\,\mathrm{d}x\,\mathrm{d}y.$$

Show further that the left-hand side is

$$2\epsilon\int_0^{2\pi}\int_0^{\infty} \frac{ru\,\mathrm{d}r\,\mathrm{d}\theta}{(r^2+\epsilon)^2}.$$

Now let $\epsilon \downarrow 0$ and either integrate by parts in r or assume the main contribution to the integral comes from near $r = 0$. Show that its value is $2\pi u(0,0)$ to lowest order, and deduce that, if $G = (1/2\pi)\log r$, then

$$\int_{\infty}^{\infty}\int_{-\infty}^{\infty} u\nabla^2 G\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G\nabla^2 u\,\mathrm{d}x\,\mathrm{d}y = u(0,0).$$

5.10. If $\nabla^2 u = 0$ in $r^2 = x^2 + y^2 \leqslant 1$, and $u(\cos\theta, \sin\theta) = g(\theta)$, separate the variables to show that

$$u = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n r^n \cos n\theta + b_n r^n \sin n\theta\right),$$

where a_n and b_n are the Fourier coefficients of g. Show that this formula can also be derived from (5.59).

5.11. Show that, if $\nabla^2 u = 0$ in a rectangle, and u is a given smooth bounded function on one side of the rectangle and zero on the other sides, then the bounded solution can be written as an explicit Fourier series. Deduce that the same is true for arbitrary smooth Dirichlet data on all four sides of the rectangle.

5.12. Show that, if $\nabla^2 u(r,\theta) = 0$ in $0 \leqslant r \leqslant 1$, $0 \leqslant \theta \leqslant \alpha$, with

$$u = \begin{cases} r^2, & \theta = \alpha,\ 0 < \alpha < \pi/2, \\ 0, & \theta = 0, \\ \sin 2\theta, & r = 1, \end{cases}$$

then

$$u = r^2\frac{\sin 2\theta}{\sin 2\alpha}.$$

Show further that if $\alpha = \pi/2$ then

$$u = -\frac{2r^2}{\pi}\left(\sin(2\theta)\log r + \theta\cos(2\theta)\right).$$

* 5.13. (i) Show that

$$\int_0^{2\pi} \frac{1 - r\cos(\theta-\alpha)}{1 + r^2 - 2r\cos(\theta-\alpha)} g(\theta)\,\mathrm{d}\theta \to \frac{1}{2}\int_0^{2\pi} g(\theta)\,\mathrm{d}\theta + \pi g(\alpha)$$

as $r \uparrow 1$ by splitting the integral into

$$\int_0^{\alpha-\epsilon} + \int_{\alpha-\epsilon}^{\alpha+\epsilon} + \int_{\alpha+\epsilon}^{2\pi},$$

where $r = 1 - \delta$ and $\delta \ll \epsilon$.

(ii) Now suppose that $u(r,\theta)$ satisfies $\nabla^2 u = 0$ in $r^2 = x^2 + y^2 < 1$, with

$$\frac{\partial u}{\partial r} + \gamma u = g(\theta) \quad \text{on } r = 1.$$

Take $G = (1/2\pi)\log|\mathbf{x} - \boldsymbol{\xi}|$ in (5.52) to show that

$$u(\boldsymbol{\xi}) = \int_0^{2\pi} \left(u\frac{\partial G}{\partial r} - G\frac{\partial u}{\partial r} \right)_{r=1} \mathrm{d}\theta,$$

and hence that, on $r = 1$, u satisfies the equation

$$u(1,\alpha) = -\frac{1}{\pi}\int_0^{2\pi} g(\theta)\log\left(\left|2\sin\left(\frac{\theta-\alpha}{2}\right)\right|\right)\mathrm{d}\theta$$
$$+ \frac{1}{\pi}\int_0^{2\pi} u(1,\theta)\left(\gamma\log\left(\left|2\sin\left(\frac{\theta-\alpha}{2}\right)\right|\right) + \frac{1}{2}\right)\mathrm{d}\theta.$$

5.14. (i) Suppose that the function $u(r,\theta)$ is written as a double Fourier inverse,

$$u(r,\theta) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{u}(k_1,k_2)\mathrm{e}^{-ir(k_1\cos\theta + k_2\sin\theta)}\,\mathrm{d}k_1\,\mathrm{d}k_2.$$

Show that u can satisfy the Helmholtz equation $\nabla^2 u + u = 0$ if $\hat{u} = 2\pi\delta(\rho - 1)f(\phi)$, where $\rho^2 = k_1^2 + k_2^2$, $\tan\phi = k_2/k_1$ and f is arbitrary. Hence derive the ***Sommerfeld representation*** for the solution of the Helmholtz equation in the form

$$u(r,\theta) = \int_0^{2\pi} \mathrm{e}^{-ir\cos(\theta-\phi)} f(\phi)\,\mathrm{d}\phi.$$

(ii) The Kontorovich-Lebedev inversion formula (5.106) on p. 193 suggests that

$$u(r,\theta) = \int_{-\infty}^{\infty} J_{i\kappa}(r)\mathrm{e}^{-i\kappa\theta} g(\kappa)\,\mathrm{d}\kappa$$

is also a general solution of the Helmholtz equation. Verify that this formula can be deduced from the Sommerfeld representation above by showing that its formal Fourier transform with respect to θ,

$$\int_{-\infty}^{\infty}\int_0^{2\pi} \mathrm{e}^{ik\theta - ir\cos(\theta-\phi)} f(\phi)\,\mathrm{d}\theta\,\mathrm{d}\phi = \int_{-\infty}^{\infty} J_{ik}(r)\mathrm{e}^{ik\phi} f(\phi)\,\mathrm{d}\phi$$
$$= J_{ik}(r)g(k), \quad \text{say.}$$

5.15. Show that, if the real symmetric matrix $\mathbf{A}$ has real eigenvalues λ_i and orthogonal eigenvectors $\mathbf{x}_i$, so that $\mathbf{x}_i^{\mathsf{T}}\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i^{\mathsf{T}}\mathbf{x}_i$, then, for any vector $\mathbf{y} = \sum c_i \mathbf{x}_i$, the smallest eigenvalue λ_0 satisfies

$$\lambda_0 \leqslant \frac{\mathbf{y}^{\mathsf{T}}\mathbf{A}\mathbf{y}}{\mathbf{y}^{\mathsf{T}}\mathbf{y}}.$$

Show that the eigenfunctions ϕ and eigenvalues $-\lambda$ of the problem

$$\nabla^2\phi + \lambda\phi = 0 \quad \text{in a region } D,$$

with

$$\frac{\partial\phi}{\partial n} + \alpha\phi = 0 \quad \text{on } \partial D,$$

where $\partial/\partial n$ is the outward normal derivative, satisfy

$$\lambda \int_D \phi^2 \, \mathrm{d}\mathbf{x} = \int_D |\nabla\phi|^2 \, \mathrm{d}\mathbf{x} + \alpha \int_{\partial D} \phi^2 \, \mathrm{d}s.$$

Deduce, as in the matrix case, that the smallest, or *principal*, eigenvalue satisfies

$$\lambda_0 \leqslant \left(\int_D |\nabla v|^2 \, \mathrm{d}\mathbf{x} + \alpha \int_{\partial D} v^2 \, \mathrm{d}s \right) \Big/ \int_D v^2 \, \mathrm{d}\mathbf{x}$$

for any smooth v.

5.16. Suppose that the eigenvalues λ_n and normalised eigenfunctions $\phi_n(\mathbf{x})$ of the Dirichlet problem for the Laplacian in a domain D are known: that is,

$$\nabla^2\phi_n = \lambda_n\phi_n \quad \text{in } D \qquad \text{with} \qquad \phi_n = 0 \quad \text{on } \partial D.$$

Multiply by ϕ_n and use Green's theorem to show that $\lambda_n < 0$. Show further that the Green's function $G(\mathbf{x}, \boldsymbol{\xi})$ for the modified Helmholtz equation, which satisfies

$$\nabla^2 G - k^2 G = \delta(\mathbf{x} - \boldsymbol{\xi}) \quad \text{in } D \qquad \text{with} \qquad G = 0 \quad \text{on } \partial D,$$

has the expansion

$$G(\mathbf{x}, \boldsymbol{\xi}) = \sum_{n=0}^{\infty} \frac{\phi_n(\mathbf{x})\phi_n(\boldsymbol{\xi})}{\lambda_n - k^2}.$$

Show that the expansion is also valid for the Helmholtz equation provided that $-k^2$ is not equal to any of the λ_n. When $-k^2 = \lambda_n$ for some n, show how to construct a modified Green's function using the arguments of p. 173, by solving

$$\left(\nabla^2 - \lambda_n\right) G(\mathbf{x}, \boldsymbol{\xi}) = \delta(\mathbf{x} - \boldsymbol{\xi}) + c\phi_n(\mathbf{x})\phi_n(\boldsymbol{\xi})$$

via an eigenfunction expansion with a suitable choice of c.

5.17. (i) Suppose that, in the notation of (5.58),

$$\nabla^2 u = 0 \quad \text{in } r > a,$$

with

$$\frac{\partial u}{\partial r} = g(\theta) \quad \text{on } r = a.$$

Using Green's theorem, show that, as $r \to \infty$,

$$u = a \log r \int_0^{2\pi} g(\theta)\, d\theta + O(1).$$

(ii) Again using the notation of (5.58), show that, on $r = a$ with $\alpha = 0$,

$$\frac{\partial}{\partial r}(RR') = \frac{1}{|\boldsymbol{\xi}|}\left(|\boldsymbol{\xi}|^2 + a^2 - 2a|\boldsymbol{\xi}|\cos\theta\right),$$

where now $|\boldsymbol{\xi}| > a$, and hence that

$$\frac{1}{RR'}\frac{\partial}{\partial r}(RR') = \frac{1}{a}.$$

(iii) Deduce that

$$\frac{1}{2\pi}\log\left(\frac{RR'}{r}\right)$$

is a Green's function for the exterior Neumann problem.

5.18. Consider the interior Neumann problem

$$\nabla^2 u = 0 \quad \text{in } r < a,$$

with

$$\frac{\partial u}{\partial r} = g(\theta) \quad \text{on } r = a,$$

where $\int_0^{2\pi} g(\theta)\, d\theta = 0$. Show that, when $\eta = 0$ in the notation of (5.56),

$$G_M = \frac{1}{2\pi}\log\left(\frac{RR'}{r}\right) + \text{constant}.$$

Does the value of the constant matter?

5.19. From the same geometrical consideration that led to (5.58), show that the Green's function for the Dirichlet problem for Laplace's equation in a sphere of radius a is

$$G(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{4\pi}\left(\frac{1}{|\mathbf{x} - \boldsymbol{\xi}|} - \frac{a}{|\boldsymbol{\xi}||\mathbf{x} - \boldsymbol{\xi}'|}\right),$$

where $\boldsymbol{\xi}'$ is the point inverse to $\boldsymbol{\xi}$ in the sphere.

5.20. Suppose θ and θ' are two-dimensional polar coordinates centred at A and B, respectively. Show that $\theta-\theta'$ is a harmonic function in any doubly-connected region surrounding, but not including, A or B. What Dirichlet data does it satisfy on any circle through A and B? Hence find the bounded solution of $\nabla^2 u = 0$ for $r < 1$ in plane polar coordinates, with

$$u(1,\theta) = \begin{cases} 1, & 0 < \theta < \pi, \\ -1, & -\pi < \theta < 0. \end{cases}$$

*5.21. Suppose a solution of $\nabla^2 u = 0$ in $y > 0$ is such that $u(x,0) = g_D(x)$ and $\partial u/\partial y\,(x,0) = g_N(x)$. Show that the Fourier transforms of g_D and g_N satisfy

$$\hat{g}_N(k) = -|k|\hat{g}_D(k),$$

and hence that

$$g_N = -\mathrm{i}h(x) * g_D'(x), \quad g_D'(x) = \mathrm{i}h(x) * g_N(x),$$

where $*$ denotes convolution and

$$\hat{h}(k) = \begin{cases} 1, & k > 0, \\ -1, & k < 0. \end{cases}$$

Show that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\epsilon|k|}\hat{h}(k)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}k = \frac{1}{2\pi}\left(\frac{1}{\mathrm{i}x+\epsilon} + \frac{1}{\mathrm{i}x-\epsilon}\right) \quad \text{for } \epsilon > 0$$

and take the limit as $\epsilon \to 0$ to deduce that $h(x) = -\mathrm{i}/\pi x$. Show that if $g_D = 1/(1+x^2)$ then $g_N = (x^2-1)/(x^2+1)^2$, and verify (5.66) or (5.68). Confirm the results by considering the function $1/(z+\mathrm{i})$.

5.22. Show that, if f is analytic in $|z| < 1$, so that

$$f(z) = \frac{1}{2\pi \mathrm{i}}\int_{|z|=1} \frac{f(t)\,\mathrm{d}t}{t-z},$$

then, for $|z| < 1$,

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{f(\mathrm{e}^{\mathrm{i}\theta})\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta}{\mathrm{e}^{\mathrm{i}\theta}-z} \quad \text{and} \quad 0 = \int_0^{2\pi} \frac{f(\mathrm{e}^{\mathrm{i}\theta})\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta}{1-\bar{z}\mathrm{e}^{\mathrm{i}\theta}}.$$

Combine these results to give $\Re f(z) = u(x,y)$, where

$$u(x,y) = \frac{1-r^2}{2\pi}\int_0^{2\pi} \frac{u(\cos\theta, \sin\theta)\,\mathrm{d}\theta}{1+r^2-2r\cos\left(\theta-\tan^{-1}(y/x)\right)}.$$

5.23. Suppose that $u(x,y)$ satisfies

$$\nabla^2 u = 0 \quad \text{in } y > 0,$$

with $u(x,0) = u_0(x)$, where $u_0(x) \to 0$ sufficiently rapidly as $x \to \pm\infty$. Show from the Green's function representation of the solution that

$$u(x,y) = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} u_0(x + y\tan\theta)\,\mathrm{d}\theta.$$

Verify by direct differentiation that $\nabla^2 u = 0$.

5.24. Suppose that $\phi(r,z)$ satisfies

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial z^2} = 0, \quad \phi(0,z) = \phi_0(z).$$

Take Hankel transforms to show that, for $z > 0$,

$$\phi(r,z) = \int_0^\infty k f(k)\mathrm{e}^{-kz} J_0(kr)\,\mathrm{d}k,$$

where $kf(k)$ is the inverse Laplace transform of $\phi(0,z)$. Now change the order of integration in

$$\phi(r,z) = \frac{1}{2\pi\mathrm{i}}\int_0^\infty \int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} \phi(0,z)\mathrm{e}^{k\zeta-kz} J_0(kr)\,\mathrm{d}k\,\mathrm{d}\zeta$$

and use the result of Exercise 4.11 to show that

$$\phi(r,z) = \frac{1}{\pi}\int_0^\pi \phi(0, z + \mathrm{i}r\cos\theta)\,\mathrm{d}\theta.$$

5.25. Show that, if $\zeta = f(z)$ is analytic, then the Neumann data satisfied by a function on a curve in the z plane is equal to that satisfied by the function on the corresponding curve in the ζ plane multiplied by $|f'(z)|$.

5.26. Show that, if ϕ is the velocity potential in an irrotational flow and $\phi + \mathrm{i}\psi = w(z)$ describes a particular flow, then

$$w(z) + \overline{w}\left(\frac{a^2}{z}\right)$$

describes a flow in which $\nabla\phi$ is tangent to $|z| = a$.

Note. If $f(z) = u(x,y) + \mathrm{i}v(x,y)$, then

$$\overline{f}(z) = \overline{f(\bar{z})} = u(x,-y) - \mathrm{i}v(x,-y).$$

5.27. Show that, for $a > b$, (5.121) can be written as

$$\phi + i\psi = U_\infty \frac{a+b}{2}\left(\frac{e^{-i\alpha}\left(z+\sqrt{z^2-c^2}\right)}{a+b} + \frac{e^{i\alpha}\left(z-\sqrt{z^2-c^2}\right)}{a-b}\right).$$

Show further that, if there is a circulation Γ around the aerofoil, so that $\phi + i\psi = U_\infty z e^{-i\alpha} + (i\Gamma/2\pi)\log z + O(1)$ as $|z| \to \infty$, then the term

$$\frac{i\Gamma}{2\pi}\log\left(\frac{z+\sqrt{z^2-c^2}}{c}\right)$$

must be added to this formula. By considering $dw/d\zeta$, show that, in the limit $b \to 0$ and $c \to a$, $dw/dz|_{z=c}$ is finite if $\Gamma = 2\pi U_\infty c\xi_0 e^{\xi_0}\sin\alpha$.

5.28. Suppose we try to solve the biharmonic equation $\nabla^4 u = 0$ in $0 < x < 1$, $0 < y < 1$, with $u = \partial u/\partial x = 0$ on $x = 0, 1$, and u and $\partial u/\partial y$ prescribed on $y = 0, 1$. Separate the variables to show that candidate solutions are

$$u(x, y) = \left((Ax + B)\cos kx + (Cx + D)\sin kx\right)e^{\pm ky},$$

as are such functions multiplied by y, where A, B, C and D are constants, and k is a complex root of $k = \pm\sin k$. Are these functions mutually orthogonal for different k?

*5.29. Consider the crack problem of §5.9.3. Show that the boundary values of the function

$$W(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}$$

satisfy

$$\frac{\partial u}{\partial x}\left(-\frac{\partial u}{\partial y} - \tau\right) = 0$$

on the whole real axis except at $x = \pm c$, and deduce that $\Im\left((W(z) - i\tau)^2\right) = 0$ there. Use Schwarz reflection to show that $W(z)$ can be extended to an analytic function with singularities only at $z = \pm c$ and $z = \infty$. Use symmetry and the behaviour of u at infinity to show that the least singular possibility for $W(z)$ is

$$(W(z) - i\tau)^2 = \frac{-\tau^2 z^2}{z^2 - c^2},$$

and hence retrieve (5.149).[101]

*5.30. By considering the limit $R \to \infty$, show that

$$\int_{|\zeta|=R>c} \frac{\zeta}{\sqrt{\zeta^2 - c^2}}\frac{d\zeta}{\zeta - x} = 2\pi i,$$

where x is real and $-c < x < c$, and the branch cut is taken along the interval $(-c, c)$ of the real line. Deform the contour to one wrapped around

[101] We are indebted to Dr A. K. Head for this remark.

the branch cut, taking care near $\zeta = x$, to deduce the result of footnote 90 on p. 206.

$*$5.31. Suppose that $\nabla^2 u(x,y,z) = 0$ in $z > 0$, with $u(x,y,0) = g_D(x,y)$ and $u \to 0$ at infinity. Show that, if

$$\hat{u}(k_1,k_2,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y,z)\mathrm{e}^{\mathrm{i}(k_1x+k_2y)}\,\mathrm{d}x\,\mathrm{d}y,$$

and if

$$\frac{\partial u}{\partial z}(x,y,0) = g_N(x,y),$$

then

$$\hat{g}_D = -\frac{1}{\sqrt{k_1^2+k_2^2}}\hat{g}_N.$$

Deduce that

$$\frac{\partial \hat{g}_D}{\partial x} = \frac{\mathrm{i}k_1}{\sqrt{k_1^2+k_2^2}}\hat{g}_N = \hat{H}\hat{g}_N,$$

say, and also that

$$\begin{aligned} H &= -\frac{1}{4\pi^2}\frac{\partial}{\partial x}\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{k_1^2+k_2^2}}\,\mathrm{e}^{-\epsilon\sqrt{k_1^2+k_2^2}}\mathrm{e}^{-\mathrm{i}(k_1x+k_2y)}\,\mathrm{d}k_1\,\mathrm{d}k_2 \\ &= -\frac{1}{4\pi^2}\lim_{\epsilon\to 0}\int_0^{2\pi}\frac{\mathrm{d}\theta}{(\mathrm{i}\cos(\theta-\tan^{-1}(y/x))+\epsilon)^2} = \frac{x}{2\pi(x^2+y^2)^{3/2}}. \end{aligned}$$

5.32. (i) Show that, if $\nabla^2 u_\pm = 0$ in $y > 0$ and $y < 0$, and $u_\pm = g_D(x)$ on $y = 0$, where g_D is continuous and $g_D(x) \to 0$ as $|x| \to \infty$, then, on $y = 0$,

$$\left[\frac{\partial u}{\partial y}\right]_-^+ = \frac{2}{\pi}\mathrm{PV}\int_{-\infty}^{\infty}\frac{g_D(\xi)-g_D(x)}{(\xi-x)^2}\,\mathrm{d}\xi.$$

(ii) Show that, if $\nabla^2 u_\pm = 0$ in $y > 0$ and $y < 0$, and $\partial u_\pm/\partial y = g_N(x)$ on $y = 0$, where $g_N \to 0$ sufficiently fast as $|x| \to \infty$, then, on $y = 0$,

$$[u]_-^+ = \frac{1}{\pi}\int_{-\infty}^{\infty} g_N(\xi)\log|x-\xi|\,\mathrm{d}\xi.$$

5.33. Show that, if $r \ll \delta \ll 1$, the integral

$$\int_{-\infty}^{\infty}\frac{g(\xi)\,\mathrm{d}\xi}{\sqrt{(z-\xi)^2+r^2}}$$

can be approximated by

$$\int_{-\infty}^{z-\delta}\frac{g(\xi)\,\mathrm{d}\xi}{z-\xi} + g(z)\int_{z-\delta}^{z+\delta}\frac{\mathrm{d}\xi}{\sqrt{(z-\xi)^2+r^2}} + \int_{z+\delta}^{\infty}\frac{g(\xi)\,\mathrm{d}\xi}{\xi-z}.$$

Show further that the middle term is

$$2g(z)\cosh^{-1}\left(\frac{\delta}{r}\right) = 2g(z)\log\left(\frac{\delta}{2r}\right)\left(1+O\left(\frac{r^2}{\delta^2}\right)\right),$$

and deduce that the original integral tends to $-2g(z)\log r$ as $r \to 0$.

5.34. Suppose that $\phi(r, z)$ satisfies

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r}\frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{\partial z^2} = 0, \quad \phi(0, z) = \frac{z}{1+z^2}.$$

Show that

$$\phi(r, z) = \Re \frac{1}{\sqrt{r^2 + (z+\mathrm{i})^2}}.$$

Where are the singularities of ϕ?

5.35. Assume that $f(x)$ is integrable, and differentiable in $0 \leqslant x < \infty$. Integrate by parts to show that

$$\begin{aligned} I(k) &= \int_0^\infty x^{1/2} f(x) \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x \\ &= \frac{\mathrm{i}}{2k}\left(g(0)\int_0^\infty x^{-1/2}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x + \int_0^\infty x^{-1/2}\left(g(x) - g(0)\right)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x\right), \end{aligned}$$

where $g(x) = f(x) + 2xf'(x)$. Deduce that

$$I(k) = \frac{1+\mathrm{i}}{2k^{3/2}}\sqrt{\frac{\pi}{2}} f(0)\left(1 + o(1)\right)$$

as $k \to \infty$.

Note. $\int_0^\infty s^{-1/2}\mathrm{e}^{\mathrm{i}s}\,\mathrm{d}s = (1+\mathrm{i})\sqrt{\pi/2}$.

5.36. (i) Generalise the argument following (5.128) to show that, if $\nabla^2 u = 0$ in $y > 0$, with $u \to 0$ as $y \to \infty$ and $u(x, 0) = g(x)$, where $g(x) \to 0$ as $x \to \pm\infty$, then $u(x, y) = \Re\, w(z)$, where

$$w(z) = \frac{1}{\pi \mathrm{i}}\int_{-\infty}^{\infty}\frac{g(\xi)\,\mathrm{d}\xi}{\xi - z}.$$

Deduce that $w(z) = O(z^{-1})$ as $|z| \to \infty$ unless $\int_{-\infty}^{\infty} g(\xi)\,\mathrm{d}\xi = 0$.

(ii) Now suppose that $\nabla^2 v = 0$ in $y > 0$, with $v \to 0$ as $y \to \infty$ and

$$\frac{\partial v}{\partial y}(x, 0) = f(x) \quad \text{for } x < 0, \qquad v(x, 0) = g(x) \quad \text{for } x > 0,$$

where f and g vanish at $\mp\infty$, respectively. Write

$$W(z) = \frac{\partial v}{\partial x} - \mathrm{i}\frac{\partial v}{\partial y}$$

and let $w(z) = z^{1/2}W(z)$. Show that, on the real axis,

$$\Re\, w(z) = h(x) = \begin{cases} x^{1/2} g'(x), & x > 0, \\ -(-x)^{1/2} f(x), & x < 0, \end{cases}$$

and deduce that

$$W(z) = \frac{z^{-1/2}}{\pi \mathrm{i}} \int_{-\infty}^{\infty} \frac{h(\xi)\,\mathrm{d}\xi}{\xi - z}.$$

Remark. This result can easily be generalised. If v and $\partial v/\partial y$ are prescribed alternately on several intervals of the real axis, then the premultiplier $z^{1/2}$, whose function is to swap real and imaginary parts (to turn Neumann data into Dirichlet), is replaced by a suitable product with square roots at the ends of the intervals. Further, if a mixed boundary value problem of this type is given in a domain that can be mapped onto a half-plane (for example, a polygon, by the Schwarz–Christoffel map; see p. 342), then the solution can also be written down; in this case the Neumann data is multiplied by the derivative of the mapping function (see Exercise 5.25).

5.37. Suppose that $u(x, y)$ satisfies the convection–diffusion equation

$$\mathbf{v} \cdot \nabla u = \nabla^2 u$$

in a domain D exterior to a semi-infinite boundary $\Gamma : y = \pm f(x)$, $0 < x < \infty$, where $f(0) = 0$ and $f(x) > 0$ for $x > 0$. Suppose also that $\mathbf{v} = \nabla\phi$, where ϕ is given and satisfies $\nabla^2\phi = 0$, that $\mathbf{v} \to (1, 0)$ at infinity, and finally that $u = 1$ on Γ and $u \to 0$ at infinity.

Show that, after the *Boussinesq transformation*, in which ϕ and its harmonic conjugate ψ (the stream function) are used as independent variables, the problem becomes

$$\frac{\partial^2 u}{\partial \phi^2} + \frac{\partial^2 u}{\partial \psi^2} = \frac{\partial u}{\partial \phi}$$

in the (ϕ, ψ) plane with the positive real axis deleted, and

$$u = 1 \quad \text{on } \psi = 0,\ \phi > 0, \qquad u \to 0 \quad \text{at infinity.}$$

Show that $u = \operatorname{erfc} \eta$, where $(\xi + \mathrm{i}\eta)^2 = x + \mathrm{i}y$ (see Exercise 5.38 for confirmation).

*5.38. Suppose that

$$\nabla^2 u = \frac{\partial u}{\partial x}$$

in $y > 0$, with

$$u(x, 0) = 1 \quad \text{for } x > 0, \qquad \frac{\partial u}{\partial y}(x, 0) = 0 \quad \text{for } x < 0,$$

and $u \to 0$ as $x^2 + y^2 \to \infty$ except on $y = 0$, $x > 0$.

(i) Show that, if $z = x + \mathrm{i}y = (\xi + \mathrm{i}\eta)^2$, then

$$\frac{\partial^2 u}{\partial \xi^2} + \frac{\partial^2 u}{\partial \eta^2} = 2\left(\xi \frac{\partial u}{\partial \xi} - \eta \frac{\partial u}{\partial \eta}\right),$$

with $u = 1$ on $\eta = 0$ and $u \to 0$ as $\eta \to \infty$. Deduce that

$$u = \frac{2}{\sqrt{\pi}} \int_{\eta}^{\infty} e^{-s^2}\, ds,$$

where $\eta = r^{1/2} \sin \theta/2$ in polar coordinates.

(ii) Derive the same result by setting $\hat{u} = \int_{-\infty}^{\infty} u e^{ikx}\, dx$, where

$$u(x,0) = f(x) \quad \text{for } x < 0, \qquad \frac{\partial u}{\partial y}(x,0) = g(x) \quad \text{for } x > 0;$$

first write that

$$\hat{g}_+ + \sqrt{k^2 - ik}\, \hat{f}_- = \frac{\sqrt{k^2 - ik}}{(ik)_+},$$

in the notation of §5.9.4, where $\sqrt{k^2 - ik} \to |k|$ as $|k| \to \infty$ with k real, and then justify writing $\sqrt{k^2 - ik} = (\sqrt{k})_+(\sqrt{k - i})_-$. Thus show that $\hat{g}_+ = (\sqrt{i/k})_+$, from which the solution in (i) can be deduced by contour integration.

* 5.39. (i) After removing the incident field, the Sommerfeld problem (5.102) and (5.104) becomes

$$(\nabla^2 + 1)u = 0,$$

with

$$u = -e^{-ix\cos\alpha} \quad \text{on } y = 0,\ x < 0,$$

together with a radiation condition. Letting

$$u(x,0) = f(x) \quad \text{for } x > 0, \qquad \left[\frac{\partial u}{\partial y}\right]_{y=0-}^{y=0+} = g(x) \quad \text{for } x < 0,$$

show, as in Exercise 5.38, that

$$\hat{f}_+(k) + \frac{i}{(k - \cos\alpha)_-} = -\frac{\hat{g}_-(k)}{2\sqrt{k^2 - 1}},$$

where $\sqrt{k^2 - 1}$ tends to $|k|$ as $k \to \infty$ with k real. *Assuming* that the radiation condition can only be satisfied if $\sqrt{k^2 - 1}$ is defined as $\left(\sqrt{k - 1}\right)_- \left(\sqrt{k + 1}\right)_+$, show that

$$\hat{f}_+(k)\left(\sqrt{k+1}\right)_+ + i\left(\frac{\sqrt{k+1} - \sqrt{\cos\alpha + 1}}{k - \cos\alpha}\right)_+ = -\frac{\hat{g}_-(k)}{2\left(\sqrt{k-1}\right)_-} - \frac{i\sqrt{\cos\alpha + 1}}{(k - \cos\alpha)_-},$$

and hence use Liouville's theorem to show that

$$\hat{g}_-(k) = -2\mathrm{i}\sqrt{\cos\alpha + 1}\left(\frac{\sqrt{k-1}}{k-\cos\alpha}\right)_- .$$

Remark. The inversion for $\hat{g}(k)$ can be manipulated to give

$$u(r,\theta) = \frac{\mathrm{e}^{\mathrm{i}(r-\pi/4)}}{\sqrt{\pi}}\left(-\operatorname{Fr}\left(\sqrt{2r}\cos\left(\frac{\theta-\alpha}{2}\right)\right)\right.$$
$$\left. + \operatorname{Fr}\left(-\sqrt{2r}\cos\left(\frac{\theta+\alpha}{2}\right)\right)\right),$$

where

$$\operatorname{Fr}(z) = \mathrm{e}^{-\mathrm{i}z^2}\int_z^\infty \mathrm{e}^{\mathrm{i}t^2}\,\mathrm{d}t.$$

This solution can be shown to satisfy the radiation condition after (5.102).

(ii) Verify that each of the two terms in $u(r,\theta)$ satisfies Helmholtz' equation by setting $u = \mathrm{e}^{\mathrm{i}r}v$ and seeking v as a function of the variable ξ in Exercise 5.38, taking the x axis in that exercise to lie along the boundaries of the directly illuminated and reflected regions, respectively.

5.40. Show that, in the parabolic coordinates of Exercise 5.38, the Helmholtz equation with $k = 1$ becomes

$$\frac{\partial^2 u}{\partial \xi^2} + \frac{\partial^2 u}{\partial \eta^2} + 4(\xi^2 + \eta^2)u = 0.$$

Although this is not apparently a traditional separation of variables situation, show that solutions exist in which $u(\xi,\eta) = U(\xi)V(\eta)$, where

$$U'' + 4(\xi^2 + \lambda)U = 0, \quad V'' + 4(\eta^2 - \lambda) = 0,$$

and λ is the separation constant.
The functions U and V are called ***parabolic cylinder functions***. Show that, if $U(\xi) = \mathrm{e}^{\mathrm{i}\xi^2}\widetilde{U}(\xi)$, then

$$\widetilde{U}'' + 4\mathrm{i}\xi\widetilde{U}' + \lambda\widetilde{U} = 0.$$

Verify, using the ideas of p. 109, that $\widetilde{U}$ has the integral representation

$$\widetilde{U} = \int_\Gamma \mathrm{e}^{\xi t - \mathrm{i}t^2/8}t^{-1-\mathrm{i}\lambda}\,\mathrm{d}t$$

for an appropriate contour Γ.

$*$5.41. The following alternative derivation of (5.112) shows that $\nabla\wedge$ can be thought of as the adjoint of $\nabla\cdot$, and vice versa.

(i) Suppose that $\nabla \wedge \mathbf{G} = \mathbf{0}$ and $\nabla \cdot \mathbf{H} = 0$. Use the identity

$$\nabla(\mathbf{a}\cdot\mathbf{b}) = (\mathbf{a}\cdot\nabla)\,\mathbf{b} + (\mathbf{b}\cdot\nabla)\,\mathbf{a} + \mathbf{a}\wedge(\nabla\wedge\mathbf{b}) + \mathbf{b}\wedge(\nabla\wedge\mathbf{a})$$

to show that, if $\mathbf{G}$ and $\mathbf{H}$ vanish sufficiently rapidly at infinity, then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\mathbf{G}\wedge(\nabla\wedge\mathbf{H}) - (\nabla\cdot\mathbf{G})\mathbf{H})\,d\mathbf{x} = \mathbf{0}.$$

(ii) Suppose that $\nabla \wedge \mathbf{H} = \mathbf{j}$ and $\nabla \cdot \mathbf{G} = \delta(\mathbf{x}-\boldsymbol{\xi})$. Show that

$$\mathbf{H} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{j}\wedge\mathbf{G}\,d\mathbf{x}.$$

(iii) Finally, show that

$$\mathbf{G} = \nabla\left(-\frac{1}{4\pi|\mathbf{x}-\boldsymbol{\xi}|}\right)$$

is an appropriate Green's vector to deduce (5.112). Further properties of vector distributions are discussed in Exercise 9.3.

5.42. Show that the function A in Exercise 5.1 satisfies

$$\frac{\partial^4 A}{\partial \bar{z}^2\,\partial z^2} = 0$$

and write it as

$$A = \frac{1}{2}\big(\bar{z}\phi_1(z) + z\phi_2(\bar{z}) + \chi_1(z) + \chi_2(\bar{z})\big),$$

where $\phi_{1,2}$ and $\chi_{1,2}$ are analytic. Use the fact that A is real to show that $\bar{\phi}_2(z) = \phi_1(z)$ and $\bar{\chi}_2(z) = \chi_1(z)$ in the notation of Exercise 5.26, and hence that

$$A = \Re\big(\bar{z}\phi_1(z) + \chi_1(z)\big).$$

Show also that

$$\begin{aligned} \sigma_x + \sigma_y &= 4\Re\,\phi_1'(z), \\ \sigma_x - \sigma_y - 2i\tau &= 2\left(\bar{z}\phi_1''(z) + \chi_1'(z)\right). \end{aligned}$$

5.43. (i) Expand the equation $\nabla\cdot(|\nabla u|^{p-2}\nabla u) = 0$ and use the criteria of Chapter 3 to show that, if $0 < |\nabla u| < \infty$, then the equation is elliptic for $p > 1$, parabolic for $p = 1$ and hyperbolic for $p < 1$.

(ii) Derive a variational formulation for the boundary value problem

$$\nabla\cdot(|\nabla u|^{p-2}\nabla u) + 1 = 0 \quad \text{in } D \qquad \text{with} \qquad u = 0 \quad \text{on } \partial D.$$

Show that the resulting functional is neither bounded above nor below if $p < 1$.

5.44. The surface $z = u(x, y)$ passes through the closed curve of intersection of $z = f(x, y)$ and the closed cylinder $h(x, y) = 0$. Show that, if the area

$$\iint \frac{dx\,dy}{\sqrt{1 + (\partial u/\partial x)^2 + (\partial u/\partial y)^2}}$$

is minimised, then

$$\left(1 + \left(\frac{\partial u}{\partial y}\right)^2\right)\frac{\partial^2 u}{\partial x^2} - 2\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}\frac{\partial^2 u}{\partial x\,\partial y} + \left(1 + \left(\frac{\partial u}{\partial x}\right)^2\right)\frac{\partial^2 u}{\partial y^2} = 0$$

in the interior of $h(x, y) = 0$, with $u = f$ on the boundary. Show that this equation can be written as

$$\nabla \cdot \left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = 0;$$

the left-hand side of this equation is called the *mean curvature* of the surface $z = u(x, y)$, which is the sum of the two principal curvatures.

5.45. Suppose that D is a domain in the (x, y) plane with a smooth boundary. The equation

$$\nabla \cdot \left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = c$$

in D, with $\nabla u \cdot \mathbf{n} = \cos\gamma$ on ∂D, is a model for a surface of constant mean curvature formed by the vertical displacement under the action of surface tension of a fixed volume of fluid occupying a cylinder whose walls, on ∂D, are in the z direction. Show that there can be no solution if

$$\frac{p}{A} < c(1 + \sec^2\gamma)^{1/2},$$

where p is the length of the perimeter of D and A is the area of D.

5.46. (i) Use the argument of §5.3 to show that, if $\lambda > 0$ and

$$\nabla^2 u + \lambda e^u = 0 \quad \text{in } D \qquad \text{with} \qquad u = 0 \quad \text{on } \partial D, \tag{5.172}$$

then u is positive. Show also that zero is a lower solution. Assuming that there is a positive solution, use induction and the maximum principle to show that there is a smallest positive solution u_M.

(ii) Show that, if T is an orthogonal matrix and $\mathbf{x}' = T\mathbf{x}$, then the Laplacian in $\mathbf{x}$ is equal to that in $\mathbf{x}'$. Deduce that, if D is a sphere, u_M is radially symmetric.

*5.47. Show that radially-symmetric solutions of (5.172) satisfy

$$\frac{d^2u}{dr^2} + \frac{m-1}{r}\frac{du}{dr} + \lambda e^u = 0 \quad \text{for } 0 < r < 1,$$

with $u(1) = 0$ when D is the unit sphere in m dimensions. Make the change of variables $s = \log r$ and $v = u + 2s + \log \lambda$ (the motivation for this will come in §6.5) to obtain

$$\frac{d^2v}{ds^2} + (m-2)\frac{dv}{ds} + e^v = 2(m-2) \quad \text{for } s < 0,$$

where

$$v(0) = \log \lambda$$

and

$$v = 2s + u_0 + \log \lambda + o(1) \quad \text{as } s \to -\infty,$$

u_0 being the value of u at $r = 0$. When $m = 2$, show that

$$\frac{ds}{dv} = \pm\frac{1}{\sqrt{2(2-e^v)}},$$

with $s = 0$ when $v = \log\lambda$, $s \to v/2$ as $v \to -\infty$ and, by integrating this explicitly or by considering the graph of $s(v)$, show that there are no solutions or two solutions according as $\lambda > 2$ or $\lambda < 2$, respectively.
Show that, for $m \geqslant 3$. the phase-plane of $(v, dv/ds)$ has one critical point at $(\log 2(m-1), 0)$, and that this is a stable spiral point (focus) for $m \leqslant 9$. Suppose that $v \leqslant v_0$ on the spiral in which $dv/ds \to 2$ as $v \to -\infty$. Show that there is only a solution if $\log\lambda < v_0$ and that there are infinitely many solutions when $\lambda = 2(m-2)$.

5.48. Show that radially-symmetric solutions in m dimensions of

$$\nabla^2 u + \lambda u^p = 0 \quad \text{for } p > 1,\ \lambda > 0$$

satisfy $u = r^{2/(1-p)}\lambda^{1/(1-p)}v$, $r = e^s$, where

$$\frac{d^2v}{ds^2} + \left(m - 2 + \frac{4}{1-p}\right)\frac{dv}{ds} + v^p + \frac{2}{1-p}\left(m - 2 + \frac{2}{1-p}\right)v = 0.$$

If $u = 1$ on $r = 1$. show that

$$v = \lambda^{-1/(1-p)} \quad \text{at } s = 0, \qquad v \to \text{constant} \cdot e^{-2s/(1-p)} \quad \text{as } s \to -\infty.$$

Show that this equation can be integrated when p is equal to the critical Sobolev exponent, $p = (m+2)/(m-2)$, to give

$$\left(\frac{dv}{ds}\right)^2 + \left(\frac{m-2}{2}\right)v^{2m/(m-2)} - \left(\frac{m-2}{2}\right)^2 v^2 = 0.$$

Deduce that in this case λ cannot exceed v_0^{p-1}, where v_0 is the maximum value of v on the closed trajectory through the origin in the $(v, dv/ds)$ plane.

5.49. Consider the problem of bifurcation from the zero solution of the equation

$$\nabla^2 u + \lambda(u - \alpha u^2) = 0 \quad \text{in } D \qquad \text{with} \qquad u = 0 \quad \text{on } \partial D,$$

at the principal eigenvalue λ_0. Suppose ϕ_0 is a corresponding positive eigenfunction. Show that ϕ_0 is a lower solution if $\lambda > \lambda_0$ and that the maximum value of ϕ_0 does not exceed $(\lambda - \lambda_0)/\alpha\lambda$. Deduce that there is a solution that is strictly positive when $\lambda > \lambda_0$, so that the bifurcation is either transcritical or what is called 'supercritical'.

*5.50. Verify that $w = Mr^{2/(1-p)}$ is a solution of

$$\nabla^2 w = w^p \quad \text{for } 0 < p < 1,$$

in the unit sphere in m dimensions, with value $M > 0$ on the boundary as long as

$$M^{p-1} = \frac{2}{1-p}\left(m - 2 + \frac{2}{1-p}\right).$$

Now suppose that u is an upper solution in which $\nabla^2 u = u^p$ in D, with $u = \overline{M} > M$ on ∂D, where D encloses the unit sphere. Making suitable assumptions about uniqueness and the applicability of the maximum principle, show that $u = 0$ at the centre of the sphere. By adjusting the position of the sphere slightly, show that u vanishes in a subregion of D, i.e. that there is a 'dead core'.

5.51. (*Jensen's inequality*) Suppose that $f'' \geqslant 0$, so that, for any X and Y,

$$f(Y) \geqslant f(X) + (Y - X)f'(X).$$

Let $w(\mathbf{x}) \geqslant 0$ and $\int_D w(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. Show that, for any X and any smooth $u(\mathbf{x})$,

$$\int_D w(\mathbf{x}) f(u(\mathbf{x}))\,\mathrm{d}\mathbf{x} \geqslant \int_D w(\mathbf{x})\big(f(X) + (u(\mathbf{x}) - X)f'(X)\big)\,\mathrm{d}\mathbf{x}$$

and hence that

$$\int_D w(\mathbf{x}) f(u(\mathbf{x}))\,\mathrm{d}\mathbf{x} \geqslant f\left(\int_D w(\mathbf{x})u(\mathbf{x})\,\mathrm{d}\mathbf{x}\right).$$

*5.52. A small massive ball rests at $x = \xi$, $y = \eta$ on a membrane stretched horizontally over a wire ∂D in the (x, y) plane; the point (ξ, η) is to be found. Show that the vertical displacement $z = u(x, y)$ approximately satisfies

$$\nabla^2 u = \frac{W}{T_0}\delta(x - \xi)\delta(y - \eta) \qquad \text{with} \qquad u = 0 \quad \text{on } \partial D,$$

where W is the weight of the ball and T_0 is the tension in the membrane. Show further that, if the ball and membrane are both smooth, then $\xi + \mathrm{i}\eta = \zeta$, where, in the notation of (5.171),

$$\left.\frac{\mathrm{d}T}{\mathrm{d}z}\right|_{z=\zeta} = 0.$$

6

Parabolic equations

Preamble

As we have already mentioned, and as this chapter emphasises, parabolic equations probably occur more commonly than any other type of partial differential equation in applied science. However, that is not the only reason why this chapter is one of the longest of the book. It also reflects the vast amount of knowledge that has been gained about many parabolic problems over recent decades. This has come about as the result of the stimulus of the many practical applications combined with the applicability of a wide variety of mathematical techniques.

We have a difficult task to present all this material in an accessible form, simply because there are so many model problems and techniques. The only attributes of parabolicity that transcend the whole chapter are the nearly universal smoothing property of parabolic operators and the difference between forward and backward parabolic equations. The techniques we use will extend beyond the Riemann–Green and Fourier methods that were applied to hyperbolic and elliptic problems; although we will not use complex variable theory other than to evaluate integrals, we will describe a battery of other methods such as maximum principles, comparison theorems, energy methods and group invariance.

In this chapter we follow convention and denote the diffusion coefficient by D; hence all spatial domains are denoted by Ω.

6.1 Linear models of diffusion

6.1.1 Heat and mass transfer

Parabolic equations arise as models of many physical processes. The most basic is the heat equation which describes the flow of heat by conduction through a stationary, homogeneous, isotropic material. This is the time-dependent version of the problem introduced in §5.1.3, so that the temperature T is no longer steady and the heat content per unit volume, ρcT, changes with time; here ρ is the density and c is the specific heat. Again taking Fourier's law of heat conduction, so that the heat flux is $\mathbf{q} = -k\nabla T$, conservation of heat in the presence of a heat source per unit volume f now gives

$$\rho c\frac{\partial T}{\partial t} = k\nabla^2 T + f, \tag{6.1}$$

assuming that ρ, c and the thermal conductivity k are all positive constants.

In the absence of any heat source, (6.1) reduces to

$$\frac{\partial T}{\partial t} = D\nabla^2 T, \tag{6.2}$$

where $D = k/\rho c$ is known as the ***thermal diffusivity***, and we will refer to both (6.1) and (6.2) as ***heat equations***. In both (6.1) and (6.2) we would expect to know the distribution of T at $t = 0$ in addition to boundary conditions imposed on spatial boundaries (possibly moving with finite speed) of the type described in §5.1.3.

These equations are often also referred to as ***diffusion equations*** because they also model diffusive mass transfer. Suppose one material, say strawberry sauce, is free to move under the sole action of molecular diffusion through another, stationary, material, say yoghurt, and suppose the concentration $c(\mathbf{x}, t)$, defined to be the ratio of the masses of sauce and yoghurt per unit total volume, is low. Then ***Fick's law*** relates the mass flux of the mobile phase to c by $\mathbf{q} = -D\nabla c$, where D is called the ***diffusivity***. Conservation of mass implies that

$$\frac{\partial c}{\partial t} = -\nabla \cdot \mathbf{q} = D\nabla^2 c$$

for constant diffusivity. If, however, the medium is moving with velocity $\mathbf{v}$ there is also mass transfer by convection. We can model this either by generalising the argument leading to (6.1) or by simply realising that the total mass flux is $\mathbf{q} = c\mathbf{v} - D\nabla c$, so that c satisfies the *convection-diffusion equation*

$$\frac{\partial c}{\partial t} = D\nabla^2 c - \nabla \cdot (c\mathbf{v}),$$

which, if $\nabla \cdot \mathbf{v} = 0$, can be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right) c = D\nabla^2 c. \tag{6.3}$$

A version of (6.3) arises when modelling the distribution c of a pollutant in a river along the x axis. With a unidirectional flow $\mathbf{v} = (v(y, z), 0, 0)$, we would write

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}\right), \tag{6.4}$$

but it now probably makes physical sense to interpret D as representing the effects of turbulence rather than molecular diffusion. The term $\partial^2 c/\partial x^2$ in (6.4) is often ignored when the river is long, and in such cases the boundary conditions at the downstream end of the river are ignored. When this approximation is made and when we restrict ourselves to steady states, c satisfies

$$v\frac{\partial c}{\partial x} = D\left(\frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}\right), \tag{6.5}$$

which shows that time does not have to be the independent variable on the left-hand side of a diffusion equation; in steady convection, 'distance downstream' is synonymous with 'time' in conventional heat conduction.

6.1.2 Probability and finance

Diffusion equations can also be obtained from 'random walk' or 'Brownian motion' models. A very simple derivation of the one-dimensional equation is as follows; it is similar to the proof-reading model of §1.1. Suppose that, at a time t, some particles occupy the lattice sites $x = 0, \pm k, \ldots$, and that the concentration $c(x,t)$ is defined to be the expected number of particles at the site at x at time t. Over the next time step, say of length h, any one particle can move to the right or left, both with probability p, or remain at its present position, with probability $1-2p$. The new expected number at x is

$$c(x,t+h) = pc(x-k,t) + (1-2p)c(x,t) + pc(x+k,t), \tag{6.6}$$

so that

$$c(x,t+h) - c(x,t) = p\big(c(x+k,t) - 2c(x,t) + c(x-k,t)\big). \tag{6.7}$$

Taking the step size and lattice separation to be small, and expanding in Taylor series about (x,t), with $k^2/h = D/p$, we recover the heat equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}. \tag{6.8}$$

This is reassuring because heat conduction comes about through the random agitation of certain modes of oscillation of atoms. If the probabilities of moving left and right had been different from each other, there would have been a drift term proportional to $\partial c/\partial x$, as in the one-dimensional form of (6.3).

In a similar vein, financial modelling also gives rise to parabolic equations. Suppose we consider an ***option***, which is a contract giving its holder the right (but not the obligation) to buy (or sell) some asset, such as a number of stock-market shares, at some specified time, say T, when the *exercise price*, a previously agreed sum of money E, is paid for the asset. Suppose the underlying asset is a share which is expected to gain in value in $0 < t < T$, but whose price is subject to unpredictable fluctuations. Suppose we buy an option instead of the share; we can gain if the share rises but, on the other hand, we may lose all our money if the share falls, since there is no point in paying E for something which costs less than that in the market. However, we can 'hedge' the option position by setting up a 'portfolio' of the option and a certain number of shares, trying to use the share holding to protect ourselves against unpredictability. As we now see, this process allows us to calculate the value $V(S,t)$ of the option to buy a share at time T as a function of the current time t and the asset value S. We suppose we have a cash balance M, and we hold a number Δ, which may vary in time, of the assets. Thus, having bought one option, the portfolio value is $P = M + S\Delta + V$. The cash balance accrues interest at a rate r, say; it also changes when we buy or sell assets and so, in a short time $\mathrm{d}t$, we receive $rM\,\mathrm{d}t$ in interest and spend $-S\,\mathrm{d}\Delta$ on assets. In the same time, the asset price changes by $\mathrm{d}S$ and the option value by $\mathrm{d}V$, so the overall change in the portfolio is

$$\mathrm{d}P = rM\,\mathrm{d}t - S\,\mathrm{d}\Delta + S\,\mathrm{d}\Delta + \Delta\,\mathrm{d}S + \mathrm{d}V = rM\,\mathrm{d}t + \Delta\,\mathrm{d}S + \mathrm{d}V.$$

We now make three important modelling assumptions. The first is that the instantaneous 'rate of return' on the asset varies randomly, so that

$$\frac{dS}{S} = \mu\,dt + \sigma\,dX, \tag{6.9}$$

where μ is a deterministic 'growth rate' for the asset; more importantly, dX is a small normal random variable[102] of mean zero and variance dt which models the uncertain response of the share price to the arrival of new information, and σ is a parameter which measures how 'volatile' the share price is. We can estimate the change in $V = V(S, t)$ in a time interval dt by writing

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2 + \cdots .$$

Now dS is given by (6.9), and we take the second bold step of assuming that the largest contribution to dS^2 is $\sigma^2 S^2\,dX^2$ and then replacing dX^2 by dt, since X has zero mean and variance dt. This second step is equivalent to the assumption that is needed to go from (6.7) to (6.8), namely $k^2 = O(h)$, and it can be described systematically using *Itô's lemma* from stochastic calculus [31]. The upshot is that

$$dP = rM\,dt + \Delta\,dS + \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}dt + o(dt).$$

This enables us to make the key observation that we can instantaneously remove *all* the randomness, represented by dS, from our portfolio by choosing Δ to be $-\partial V/\partial S$.

The final step is to use the idea of *no arbitrage*, which is the technical term for the non-existence of a 'free lunch'. In this context, it means that it is impossible to earn more than the risk-free interest rate r for a risk-free portfolio, so $dP = rP\,dt$ and hence we derive the *Black–Scholes equation*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} = r\left(V - S\frac{\partial V}{\partial S}\right) \tag{6.10}$$

for the value of our option. Note that, as compared to the heat equation, the Black–Scholes equation is a 'backward' parabolic equation with a convective term $rS\,\partial V/\partial S$ and a 'source' term rV, and we expect intuitively that a final condition should be imposed. This is that V equals $S - E$ at $t = T$, provided that $S > E$, which represents the proceeds of exercising the option and immediately selling the asset; if, on the other hand, $S < E$, the holder will not wish to pay more for the asset than it is worth, and in this case the option is not exercised, so $V = 0$. We note that if S vanishes at some time it remains zero, according to (6.9), so the boundary condition $V = 0$ at $S = 0$ must also hold.[103]

[102]There have been many more large swings in world markets than would be consistent with the assumption of a normal random variable; this model must be used with caution!

[103]The discussion we have given is just the starting point of an enormous range of valuation problems for 'vanilla' and 'exotic' options in the Black–Scholes framework; most of them are boundary value problems for the Black–Scholes equation or a variant of it, and we refer the reader to [47].

6.1.3 Electromagnetism

The heat equation also appears in vector form in electromagnetism. We recall from §4.7.2 that, in a suitable system of units and with no net charge, the electric field $\mathbf{E}$, magnetic field $\mathbf{H}$ and current density $\mathbf{j}$ satisfy

$$-\mu\frac{\partial \mathbf{H}}{\partial t} = \nabla \times \mathbf{E}, \quad \nabla \times \mathbf{H} = \mathbf{j} + \epsilon\frac{\partial \mathbf{E}}{\partial t}, \quad \nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{H} = 0.$$

Suppose that we are now considering situations where the time scale is much longer than that for electromagnetic wave propagation, which here means that ϵ is small; the second equation then becomes $\mathbf{j} = \nabla \times \mathbf{H}$ to lowest order. Suppose also that current-carrying material is present, which means that to close the model there must be a law relating $\mathbf{j}$ and $\mathbf{E}$ in the material. For many materials this is Ohm's law, $\mathbf{j} = \sigma\mathbf{E}$, where σ is the electrical conductivity, assumed constant. Combining the equations, and using the fact that $\nabla \wedge (\nabla \wedge \mathbf{E}) = -\nabla^2\mathbf{E}$ since $\nabla \cdot \mathbf{E} = 0$, yields the vector diffusion equation

$$\frac{\partial \mathbf{j}}{\partial t} = \frac{1}{\mu\sigma}\nabla^2\mathbf{j}, \tag{6.11}$$

just as long as σ is a constant. This equation is also satisfied by $\mathbf{E}$ and $\mathbf{H}$, and is called the *equation of eddy currents*.

6.1.4 General remarks

The prevalence of models of the form

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = D\nabla^2 u + f \tag{6.12}$$

suggests that we start our analysis of parabolic equations by studying linear problems of this type, where f depends on u at most linearly, while $\mathbf{v}$ and D are independent of u. The independent variables $\mathbf{x}$ and t are usually identified with space and time, respectively. The variable t may be increasing ($dt > 0$) or decreasing ($dt < 0$), and we will soon see the vital role of the sign of $D\,dt$. In the examples above this product is positive for models being used to predict the future via a forward equation. The only backward equation is the Black–Scholes equation, which models the assimilation of information, rather than its loss, as the expiry date approaches; at expiry, the option value is known with certainty. Note that, if D is constant, then it can be taken to be ± 1 by a suitable change of variables that does not involve time reversal. Also, concerning nomenclature, (6.12) is often called a convection-diffusion or *reaction-diffusion equation*, depending whether $f = 0$ or $\mathbf{v} = \mathbf{0}$, respectively.

However, before we can start work on (6.12), we need to consider the vital questions of what are likely to be the appropriate initial and boundary conditions.

6.2 Initial and boundary conditions

As shown in Chapter 3, the characteristics of (6.12) in one space dimension are given by $dt^2 = 0$, so that each line t = constant is a double characteristic (in more

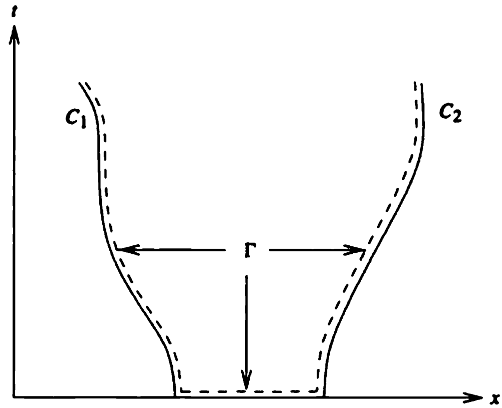


Fig. 6.1 Boundary curves $C_{1,2}$ and the parabolic boundary Γ.

dimensions, t = constant are characteristic surfaces). Information thus propagates at infinite speed along the characteristics. Moreover, it is easy to see that the relation (3.21) that has to hold along characteristics is automatically satisfied when (3.20) is true. Hence we can expect to be able to prescribe initial data $u = g(\mathbf{x})$ at $t = 0$, which is often the case in practical applications. Additionally, unless the equation holds in the whole space, boundary data must also be imposed.[104] As in §5.1.3, the only linear boundary conditions are Dirichlet, Neumann or Robin, with u, the outward normal derivative $\partial u/\partial n$, or the combination $\partial u/\partial n + \alpha u$ being specified, respectively; in the last case,[105] we must remember our earlier strictures about the sign of α on p. 164.

For problems with only one space dimension, the boundary conditions are imposed on two curves $C_{1,2}$ in the (x, t) plane, as in Fig. 6.1. So as not to be parallel to characteristics, these curves must be nowhere perpendicular to the t axis. Thus they have finite speed, and they are usually constant in time, i.e. parallel to the t axis, so that the characteristics intersect each such boundary once. Moreover, since we expect the temperature in a heat conducting material to be determined just by its initial value and what happens on its boundary, we also expect only one boundary condition on the 'time-like' boundaries $C_{1,2}$. For problems with more than one space dimension, the elliptic combination of second-order derivatives in (6.12) together with our discussion in §5.2 suggests the appropriateness of imposing one condition on a closed time-like boundary rather than two in some places and none in others.

The relevance of Fig. 6.1 to the physical interpretation of the models derived above suggests that we refer to the conditions imposed at $t = 0$ as *initial conditions* for problems to be solved in $t > 0$, so that $t = 0$ is called 'space-like'. Remember

[104]For problems in unbounded domains, growth conditions at infinity are needed to ensure uniqueness, as we shall see in §6.4.2.

[105]We take $\beta = 1$ in our earlier version of the Robin boundary condition.

that the first example of §3.1 confirms that one initial condition should be given rather than one 'final' one. This is also apparent from considering the limiting procedure of obtaining the heat equation from the hyperbolic equation

$$\epsilon^2 \frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

as $\epsilon \to 0$, where the argument of §3.4.3 also suggests that when $D\,\mathrm{d}t > 0$ only one initial condition should be imposed at $t = 0$.

There are many methods for proving existence and uniqueness results for parabolic equations, most of which are too abstract for this book. As in the previous two chapters, we will mostly be concerned, especially in §§6.4 and 6.5, with explicit representations of solutions, so that the existence question will not arise. However, the maximum principle is such a generally applicable tool for proving uniqueness that we will begin by describing its use in §6.3, before going on to describe techniques used to obtain explicit solutions to linear equations in integral and series form. Another way of obtaining special solutions that has only received brief attention in Chapters 4 and 5 is that of reducing the number of independent variables through invariance properties. We consider this in some detail in §6.5, before applying it to nonlinear equations in §6.6, where other methods suitable for studying such equations are also reviewed. The final section takes a quick look at some second-order parabolic systems and higher-order parabolic equations.

6.3 Maximum principles and well-posedness

Even more than in the elliptic case, the idea of maximum principles is invaluable in assessing well-posedness properties of parabolic equations. Parabolic maximum principles, like those of the preceding chapter, apply in any number of space dimensions. The maximum principle for the simplest inhomogeneous heat equation states that, if

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(x, t)$$

and $f \leqslant 0$, then u takes its maximum on the *parabolic boundary*, denoted by Γ, where the initial and boundary conditions are given (see Fig. 6.1). Thus, when the equation holds for x in Ω and $0 < t < T$, u is largest either on $\partial\Omega$ or at $t = 0$. This is of course what we expect physically. If heat is being lost through a volumetric heat sink with $f \leqslant 0$, then heat flows into Ω from the boundary and is removed in the interior, and the maximum temperature must be on the boundary if it is not at $t = 0$; equally, with $f \geqslant 0$, u takes its minimum on the parabolic boundary.

The proof is similar to that for the maximum principle for Laplace's equation in §5.3. Consider first $f < 0$; then, at an interior maximum $\partial u/\partial t = 0, \nabla u = \mathbf{0}$ and $\nabla^2 u \leqslant 0$, leading to a contradiction. For a maximum on $t = T$, $\partial u/\partial t \geqslant 0$, $\nabla u = \mathbf{0}$ and $\nabla^2 u \leqslant 0$, again leading to a contradiction, so the maximum must lie on Γ. To consider $f \leqslant 0$, write $v = u + \epsilon|\mathbf{x}|^2/2m$, taking m to be the number of space dimensions. Then

$$\frac{\partial v}{\partial t} - \nabla^2 v = f - \epsilon < 0,$$

and hence v takes its maximum on Γ. Thus $u \leqslant v \leqslant \max_\Gamma v \leqslant \epsilon a^2/2m + M$, where M is the maximum value of u on Γ and a is the largest value of $|\mathbf{x}|$. Taking the limit as $\epsilon \to 0$, we see that $u \leqslant M$, so that u has its maximum value on Γ; note that, in this case, u can also take this value in the interior of Ω when $t > 0$.

It can be shown that the maximum principle also holds for the more general equation

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = D\nabla^2 u + f(\mathbf{x}, t),$$

as long as D is positive and $\mathbf{v}$ is bounded, and also when a more general elliptic operator

$$\mathcal{L}u = \sum_{i,j} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial}{\partial x_j}\right)$$

replaces the Laplacian, (a_{ij}) being a positive definite matrix. However, more care must be taken with the slightly modified equation $\partial u/\partial t = \nabla^2 u + au$. Consider, for example, the one-dimensional case when $u = 0$ on $x = 0, 1$ and $u = \sin \pi x$ at $t = 0$, $0 < x < 1$. The maximum value of u on the parabolic boundary is unity, but the exact solution is $u = e^{(a-\pi^2)t} \sin \pi x$, whose maximum in $t > 0$ exceeds unity when $a > \pi^2$.

The maximum principle is the key tool for proving the fundamental uniqueness theorem for the Dirichlet problem

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = \nabla^2 u + f(\mathbf{x}, t) \quad \text{for } \mathbf{x} \text{ in } \Omega,\ t > 0, \tag{6.13}$$

$$u(\mathbf{x}, 0) = g(\mathbf{x}) \quad \text{for } \mathbf{x} \text{ in } \Omega, \tag{6.14}$$

$$u = h(\mathbf{x}) \quad \text{for } \mathbf{x} \text{ on } \partial\Omega,\ t > 0. \tag{6.15}$$

This follows because the difference of any two solutions satisfies (6.13)–(6.15) with f, g and h replaced by zero, and, since this difference must take its maximum and minimum values on Γ, it must be zero everywhere. Also, small changes in the initial and boundary data bound any consequent change in the solution so that, assuming a solution exists, the problem is also well posed.

*6.3.1 The strong maximum principle

Some uniqueness results rely on the use of the *strong maximum principle* which applies to *parabolic inequalities*. Suppose u is such that

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u \leqslant \nabla^2 u$$

for x in Ω and $0 < t < T$, with $\mathbf{v}$ a bounded function of $\mathbf{x}$ and t, and again let M be the maximum of u on Γ. Then, roughly speaking, this principle states that either u is less than M for all x in Ω and $0 \leqslant t < T$, and $\partial u/\partial n$ is strictly positive at any point on $\partial\Omega$ where $u = M$, or u is identical to M for all $\mathbf{x}$ over some interval

$0 \leqslant t \leqslant \tau < T$; a more precise statement and proof are given in [35]. This stronger result enables us to show, for example, that, if

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = \nabla^2 u \quad \text{in } \Omega,$$

with $u \geqslant 0$ at $t = 0$ and $\partial u/\partial n + \alpha u \geqslant 0$ on $\partial\Omega$, where $\alpha \geqslant 0$, then $u \geqslant 0$ for $\mathbf{x}$ in Ω, $t > 0$. This result follows from the fact that any strictly negative minimum, say m, of u is attained on $\partial\Omega$, by the maximum principle; hence, by the strong maximum principle, either $u > m$ in Ω, which implies the contradiction $0 > \partial u/\partial n = -\alpha m$ at a point of $\partial\Omega$, or $u \equiv m$ over a time interval containing $t = 0$, which is also a contradiction.

Interestingly, the sign of α is not so important when we consider the uniqueness of the solution of the Robin problem (6.13) and (6.14) with

$$\frac{\partial u}{\partial n} + \alpha u = h(\mathbf{x}) \quad \text{on } \partial\Omega. \tag{6.16}$$

To establish this uniqueness, we need the strong maximum principle. Suppose that $\alpha \geqslant 0$ and that the difference between two solutions is somewhere positive. Then the maximum difference M is taken on the boundary. Since the difference is not identically M, the normal derivative, according to the strong maximum principle, is positive, contradicting the boundary condition. Similarly, the difference cannot be negative. If, on the other hand, $\alpha < 0$, we can make a change of variable $u = g(\mathbf{x})u'$, with g chosen to be strictly positive in Ω and such that $\partial g/\partial n + \alpha g > 0$ on $\partial\Omega$; then we can use the argument above for u' and uniqueness again follows. This is in sharp contrast to the elliptic case of the Robin problem of §5.2.1, where we recall that uniqueness depended essentially on the sign of α.[106]

Note that, if we had attempted to solve any of the problems above in $t < 0$, that is 'backwards in time', then the maximum principle would have allowed a maximum to occur at $t = \tau < 0$. No uniqueness or well-posedness would follow. Indeed, ill-posedness of the backward heat equation through lack of continuous dependence on the data has already been demonstrated in §3.1 and will be encountered again and again. The situation is even worse when we come to consider the existence of solutions of backward heat equations, as we shall see in the next section.

6.4 Green's functions and transform methods for the heat equation

6.4.1 Green's functions: general remarks

In attempting to construct Riemann functions for hyperbolic equations in Chapter 4 and Green's functions for elliptic equations in Chapter 5, the quickest procedure was to use Green's theorem on the integral of $G\mathcal{L}u - u\mathcal{L}^*G$ over a suitable domain. There $\mathcal{L}$ denoted the differential operator, so that $\mathcal{L}u = f$, and $\mathcal{L}^*$ was the adjoint operator, chosen so that the integrand was in divergence form. We hope

[106] Of course, the large-time behaviour of the solution of the parabolic problem depends crucially on the sign of α.

that by now the reader has become sufficiently familiar with the generalised functions approach to this procedure that the classical approach will be unnecessary here. Hence, to construct a Green's function for the heat equation

$$\mathcal{L}u \equiv \frac{\partial u}{\partial t} - \nabla^2 u = f(\mathbf{x}, t) \tag{6.17}$$

in a region Ω with initial and Dirichlet boundary conditions $u(\mathbf{x}, 0) = g(\mathbf{x})$ in Ω and $u = h(\mathbf{x})$ on $\partial\Omega$, we require that $G(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$ satisfies

$$\mathcal{L}^*G \equiv -\frac{\partial G}{\partial t} - \nabla^2 G = \delta(\mathbf{x} - \boldsymbol{\xi})\delta(t - \tau) \quad \text{in } \Omega. \tag{6.18}$$

Furthermore, we need

$$G = 0 \quad \text{on } \partial\Omega, \tag{6.19}$$

and, remembering the analogy with the ordinary differential equation initial value problem (4.11), we set

$$G = 0 \quad \text{for } t = T > \tau. \tag{6.20}$$

Recalling that G is defined by working backwards from $t = T$, and since the delta function is non-negative, it is apparent from the maximum principle that G is positive. More importantly, from (6.17)–(6.20), we obtain

$$\int_0^T \int_\Omega \left(G\left(\frac{\partial u}{\partial t} - \nabla^2 u\right) + u\left(\frac{\partial G}{\partial t} + \nabla^2 G\right)\right) \mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_0^T \int_\Omega fG \,\mathrm{d}\mathbf{x}\,\mathrm{d}t - u(\boldsymbol{\xi}, \tau),$$

and this yields the fundamental result that

$$u(\boldsymbol{\xi}, \tau) = \int_\Omega u(\mathbf{x}, 0)G(\mathbf{x}, 0; \boldsymbol{\xi}, \tau)\,\mathrm{d}\mathbf{x} + \int_0^T \int_\Omega fG\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_0^T \int_{\partial\Omega} u\frac{\partial G}{\partial n}\,\mathrm{d}\mathbf{x}\,\mathrm{d}t. \tag{6.21}$$

As was the case for the corresponding Riemann and Green's function representations in Chapters 4 and 5, (6.21) can easily be used to prove well-posedness for parabolic equations just as long as we know that G exists and is unique. Indeed, these representations (4.14), (5.52) and (6.21) have removed the sting from our three classes of linear partial differential equations, because they display the structure of the 'inverses' of all the partial differential operators that we have encountered in these chapters. The value of those three results cannot be overestimated.

Before we set about finding G in special cases, we first make our customary general remarks about the effect of the singularity induced by the right-hand side of (6.18) at points away from $\mathbf{x} = \boldsymbol{\xi}$, $t = \tau$. First, by thinking of the limit of a hyperbolic problem when the characteristics coincide, we are led to expect some kind of discontinuity in G to propagate along the characteristic $t = \tau$, although we will have to wait until the next section to see exactly what this discontinuity is. We can, however, elucidate what happens on $t = \tau$ by the following comparison.

- Consider the effect of taking a Laplace transform of the problem

$$\frac{\partial u_1}{\partial t} - \nabla^2 u_1 = \delta(\mathbf{x})\delta(t), \tag{6.22}$$

with $u_1(\mathbf{x},t) = 0$ for $t < 0$. The answer is

$$\tilde{u}_1(\mathbf{x},p) = \int_0^\infty u_1(\mathbf{x},t)\mathrm{e}^{-pt}\,\mathrm{d}t = \lim_{\epsilon\downarrow 0}\int_{-\epsilon}^\infty u_1(\mathbf{x},t)\mathrm{e}^{-pt}\,\mathrm{d}t,$$

where $p\tilde{u}_1 - \nabla^2\tilde{u}_1 = \delta(\mathbf{x})$.

- However, if

$$\frac{\partial u_2}{\partial t} - \nabla^2 u_2 = 0 \quad \text{in } t > 0,$$

with

$$u_2 = \delta(\mathbf{x}) \quad \text{at } t = 0, \tag{6.23}$$

then $p\tilde{u}_2 - \delta(\mathbf{x}) - \nabla^2\tilde{u}_2 = 0$.

Hence $u_1 = u_2$ in $t > 0$, i.e. the effect of the right-hand side of (6.22) is equivalent to that of the boundary condition (6.23). This means that, by replacing t by $\tau - t$, we can assert that (6.18)-(6.20) is equivalent to

$$\mathcal{L}^*G = 0 \quad \text{for } 0 < t < \tau, \tag{6.24}$$

with

$$G = 0 \quad \text{on } \partial\Omega, \tag{6.25}$$

and

$$G = \delta(\mathbf{x} - \boldsymbol{\xi}) \quad \text{at } t = \tau, \tag{6.26}$$

$$G = 0 \quad \text{in } t > \tau. \tag{6.27}$$

Hence we see that G represents a 'hot spot' at $t = \tau$, and that it is a function only of $\mathbf{x}$, $\boldsymbol{\xi}$ and $\tau - t$. As a function of $\mathbf{x}$ and t, G satisfies the adjoint backward heat equation (6.18), while as a function of $\boldsymbol{\xi}$ and τ it satisfies the original forward equation (6.17). We will replace $G(\mathbf{x},t;\boldsymbol{\xi},\tau)$ by $G(\mathbf{x},\tau - t;\boldsymbol{\xi})$ in future.

6.4.2 The Green's function for the heat equation with no boundaries

This is the simplest geometry of all and its study will enable us to identify the precise form of the singularity at $\mathbf{x} = \boldsymbol{\xi}$, $t = \tau$, which we expect to be independent of Ω in any given number of space dimensions. We begin in just one dimension, and it is convenient to set $t' = \tau - t$, $x' = x - \xi$ and $G(x - \xi, \tau - t) = G'(x', t')$, to give the initial value problem

$$\begin{gathered}\frac{\partial G'}{\partial t'} = \frac{\partial^2 G'}{\partial x'^2} \quad \text{for } t' > 0, \\ G' = \delta(x') \quad \text{for } t' = 0, \qquad G' \to 0 \quad \text{as } |x'| \to \infty,\end{gathered} \tag{6.28}$$

which can be solved explicitly in a variety of ways, say by Fourier transforms. The solution describes the conduction of heat from a localised hot spot in an infinite conducting medium, and to help with this interpretation we replace x', t' and $G'(x',t')$ by x, t and $\mathcal{G}(x,t)$, respectively.

In terms of the Fourier transform $\hat{\mathcal{G}}(k,t) = \int_{-\infty}^{\infty} e^{ikx}\mathcal{G}(x,t)\,dx$, the problem becomes

$$\frac{\partial \hat{\mathcal{G}}}{\partial t} = -k^2\hat{\mathcal{G}} \quad \text{for } t > 0, \qquad \hat{\mathcal{G}}(k,0) = 1. \tag{6.29}$$

This gives

$$\hat{\mathcal{G}}(k,t) = e^{-k^2 t},$$

and, from the Fourier inversion theorem,

$$\begin{aligned}\mathcal{G}(x,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-k^2t-ikx}\,dk \\ &= \frac{1}{2\pi}e^{-x^2/4t}\int_{-\infty}^{\infty} e^{-t(k+ix/2t)^2}\,dk \\ &= \frac{1}{2\pi\sqrt{t}}e^{-x^2/4t}\int_{-\infty+ix/2\sqrt{t}}^{\infty+ix/2\sqrt{t}} e^{-s^2}\,ds,\end{aligned}$$

where we have written $k = s/\sqrt{t} - ix/2t$. The final integrand is analytic in $0 < \Im s < x/2\sqrt{t}$, and vanishes for large $\Re s$. By Cauchy's theorem, the integral is equivalent to one along the real axis. Thus, finally,

$$\mathcal{G}(x,t) = \frac{e^{-x^2/4t}}{2\sqrt{\pi t}} \quad \text{for } t > 0, \tag{6.30}$$

since $\int_{-\infty}^{\infty} e^{-s^2}\,ds = \sqrt{\pi}$; its delta function behaviour at $t = 0$ is an example of (4.12). The formula (6.30) reveals an *essential singularity* in $\mathcal{G}$ at $t = 0$, with all the time derivatives vanishing except at $x = 0$; this is the very weak singularity that propagates along the characteristic through the origin when a heat source is switched on there. We expect this singularity to be generic along the characteristic $\tau = t$ for any G satisfying (6.18).

When we revert to our original notation, in which x and t are the physical variables, the solution of (6.24)–(6.27) on the whole line in one dimension is

$$G(x,\tau - t;\xi) = \mathcal{G}(x-\xi,\tau - t) = \frac{1}{2\sqrt{\pi(\tau - t)}}e^{-(x-\xi)^2/4(\tau-t)}. \tag{6.31}$$

We can now use the statement

$$G(x,\tau - t;\xi) = \frac{1}{2\sqrt{\pi(\tau - t)}}e^{-(x-\xi)^2/4(\tau-t)} + O(1) \quad \text{as } x \to \xi,\ t \to \tau$$

to replace (6.26) in the same way that

$$G = \frac{1}{2\pi}\log|\mathbf{x} - \boldsymbol{\xi}| + O(1)$$

was used for Laplace's equation in two space dimensions in (5.49). The function (6.30) is called the *elementary* or *fundamental* solution of the heat equation and it allows us to make several remarks.[107]

When $u(x,0) = g(x)$, the Green's function representation gives

$$u(\xi,\tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} g(x) e^{-(x-\xi)^2/4\tau} \, dx, \tag{6.32}$$

which can be thought of as the result of distributing elementary solutions on $t = 0$ with density g. This solution can be seen to be analytic for $t > 0$ in both x and t, even for quite irregular initial data,[108] and it confirms that u is positive for positive initial data.

Although (6.32) gives an explicit representation for the solution, it implicitly demands that $g(x)$ does not grow too rapidly as $|x| \to \infty$. The growth condition to ensure that (6.32) does represent the unique solution is that there should exist some constant K such that $|g| = O(e^{Kx^2})$ as $|x| \to \infty$. This can be proved by considering the problem on a finite interval with zero initial data, as in Exercise 6.4. However, we can see that non-zero 'eigensolutions', which satisfy $u(x,0) = 0$, might possibly exist by noting that

$$u(x,t) = \sum_{n=0}^{\infty} f^{(n)}(t) \frac{x^{2n}}{(2n)!} \tag{6.33}$$

satisfies the heat conduction equation for all x and $t \geqslant 0$ just as long as f is infinitely differentiable and the series converges. By allowing f and all its derivatives to vanish as $t \downarrow 0$, (6.33) is such an eigensolution as long as the series converges (see Exercise 6.5). Unfortunately, it is quite difficult to extract the behaviour of (6.33) as $|x| \to \infty$.[109]

At a more elementary level, (6.30) permits us again to demonstrate explicitly the ill-posedness of the backward heat equation. Suppose that we seek u such that

$$\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{for } t > 0,$$

with $u = \sqrt{\epsilon} e^{-x^2/4\epsilon}/2\sqrt{\pi}$ at $t = 0$ and $u \to 0$ as $|x| \to \infty$, so that $u(x,0) \to 0$ as $\epsilon \to 0$ for all x. Then, from (6.30),

$$u = \frac{\epsilon e^{-x^2/4(\epsilon - t)}}{2\sqrt{\pi(\epsilon - t)}},$$

which tends to infinity at $x = 0$ as $t \to \epsilon$, and ϵ can be taken as small as we please.

[107] The implication of (6.30) for the Brownian motion model (6.8) is described in Exercise 6.6.

[108] Note that, at a jump discontinuity of g, $u(\xi,\tau)$ tends to the average of g from either side of ξ as $\tau \to 0$.

[109] It is possible to show the surprising result that the heat equation can be solved in $-\infty < x < \infty$ with arbitrary data at $t = 0$ *and* at $t = T > 0$, assuming we allow sufficient growth as $|x| \to \infty$. On the other hand, if $|u(x,t)| < e^{Kx^2}$ as $x \to \pm\infty$ for *all* $K > 0$, then the solution to the initial value problem exists and is unique for all $t > 0$.

The form of the solution (6.32) also indicates that, during the evolution of u, details of the initial data are lost and all that is remembered after a long time is some multiple of the fundamental solution. We can see this by writing

$$u(\xi,\tau) = (4\pi\tau)^{-1/2}\int_{-\infty}^{\infty} g(x)\mathrm{e}^{-y^2/4}\mathrm{e}^{-x^2/4\tau+xy/2\sqrt{\tau}}\,\mathrm{d}x,$$

where $\xi = \sqrt{\tau}y$. For large τ, this can be approximated by

$$\frac{1}{2\sqrt{\pi\tau}}\int_{-\infty}^{\infty} g(x)\mathrm{e}^{-y^2/4}\,\mathrm{d}x = \frac{g_0}{2\sqrt{\pi\tau}}\mathrm{e}^{-\xi^2/4\tau}, \tag{6.34}$$

where g_0 is the total amount of heat in the initial condition.[110]

Many of the statements above can be trivially generalised to the case $\Omega = \mathbb{R}^m$, $m > 1$. The principal result is that (6.31) becomes

$$G(\mathbf{x},\tau - t;\boldsymbol{\xi}) = \big(4\pi(\tau - t)\big)^{-m/2}\mathrm{e}^{-|\mathbf{x}-\boldsymbol{\xi}|^2/4(\tau-t)} \quad \text{for } 0 < t < \tau \tag{6.35}$$

when we revert to our original variables; (6.35) can be derived by an m-dimensional Fourier transform and exploiting the spherical symmetry in $\mathbf{x}-\boldsymbol{\xi}$, and we will derive it in another way in §6.5.

6.4.3 Boundary value problems

6.4.3.1 Green's functions and images

In principle, we can now subtract out the singular behaviour of the Green's function implied by (6.30) or (6.35) to obtain a boundary value problem for the 'regular part' of G in which all the data is well behaved. However, the problem of finding G in any particular example is just as hard as it was for elliptic equations and we again have to revert to a case-by-case enumeration. There are just two general remarks we can make first.

From the fact that the Laplace operator is self-adjoint, we expect that the spatial part of G is symmetric for appropriate boundary conditions, i.e.

$$G(\mathbf{x},\tau - t;\boldsymbol{\xi}) = G(\boldsymbol{\xi},\tau - t;\mathbf{x}),$$

but clearly there is no symmetry when we exchange t and τ. Moreover, it can be shown, by methods that will regrettably have to wait until Chapter 8, that the presence of boundaries makes G different from the form (6.35) by a function characterised by the property that it increases as the 'geodesic distance' $d(\mathbf{x},\boldsymbol{\xi})$ between $\mathbf{x}$ and $\boldsymbol{\xi}$ decreases, at least for small values of $\tau - t$; d is the shortest distance between $\mathbf{x}$ and $\boldsymbol{\xi}$, travelling along the boundary if necessary. This makes an interesting comparison with the 'boundary correction' to the Green's function for Laplace's equation, discussed in §5.12.

We conclude this section with some explicit representations for Green's functions in simple cases. A much more comprehensive catalogue can be found in [9]. We begin with two examples in one space dimension.

[110] When $g_0 = 0$, a more precise estimate shows that, in general, u is of $O(\tau^{-3/2})$ as $\tau \to \infty$.

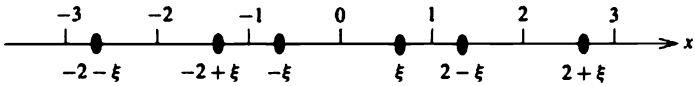


Fig. 6.2 Hot spot at $x = \xi$, and its images.

Example 6.1 (*The Neumann problem for the unit interval*) The general problem is to solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f \quad \text{for } 0 < x < 1, t > 0,$$

$$-\frac{\partial u}{\partial x} = h_0 \quad \text{for } x = 0, \qquad \frac{\partial u}{\partial x} = h_1 \quad \text{for } x = 1, \qquad u = g(x) \quad \text{at } t = 0.$$

The Green's function G must then satisfy

$$\frac{\partial G}{\partial t} + \frac{\partial^2 G}{\partial x^2} = 0 \quad \text{for } 0 < x < 1, 0 < t < \tau,$$

$$\frac{\partial G}{\partial x} = 0 \quad \text{on } x = 0 \text{ and } x = 1, \qquad G = \delta(x - \xi) \quad \text{at } t = \tau.$$

We can proceed by generalising the method of images introduced in §5.6.1.3. Recalling that $(4\pi t)^{-1/2} e^{-(x-\xi)^2/4t}$ represents the temperature evolving from a hot spot at $t = 0$, $x = \xi$, to obtain zero derivatives at $x = 0$ an image hot spot must be introduced at $x = -\xi$. Then, in order to satisfy $\partial G/\partial x = 0$ at $x = 1$, further images are needed at $x = 2 - \xi$ and $x = 2 + \xi$. Repeating, there is the 'real' hot spot at $x = \xi$, with images at $x = -\xi$, and $\dots, -4 \pm \xi, -2 \pm \xi, 2 \pm \xi, 4 \pm \xi, \dots$ (see Fig. 6.2).

Thus, the Green's function is[111]

$$G(x, \tau - t; \xi) = (4\pi(\tau - t))^{-1/2} \sum_{m=-\infty}^{\infty} \left(e^{-(x+2m-\xi)^2/4(\tau-t)} + e^{-(x+2m+\xi)^2/4(\tau-t)} \right), \tag{6.36}$$

which is clearly symmetric in x and ξ, and positive. This problem can also be approached by separating the variables, as in the next example.

Example 6.2 (*Zero Dirichlet data for the unit interval*) As with hyperbolic and elliptic equations, separation of variables can be used to find Green's functions as Fourier series. The problem for heat flow in the unit interval with zero Dirichlet data is

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } 0 < x < 1,$$

$$u = 0 \quad \text{on } x = 0 \text{ and } x = 1, \qquad u = g(x) \quad \text{at } t = 0.$$

[111] An alternative derivation is given in Exercise 6.8.

Separable solutions $u = T(t)X(x)$ are

$$u = \mathrm{e}^{-m^2\pi^2 t}\sin(m\pi x), \quad m \text{ an integer},$$

$\{\sin m\pi x\}$ being a complete orthogonal set of eigenfunctions of the self-adjoint operator $\mathrm{d}^2/\mathrm{d}x^2$ with zero Dirichlet conditions.

The general solution is then

$$u = \sum_{m=1}^{\infty} a_m \mathrm{e}^{-m^2\pi^2 t}\sin(m\pi x);$$

the Fourier coefficients a_m are determined by the initial condition as

$$a_m = 2\int_0^1 g(\xi)\sin(m\pi\xi)\,\mathrm{d}\xi.$$

The solution to the boundary value problem is then

$$u(x,t) = 2\int_0^1 g(\xi)\sum_{m=1}^{\infty}\mathrm{e}^{-m^2\pi^2 t}\sin(m\pi x)\sin(m\pi\xi)\,\mathrm{d}\xi, \tag{6.37}$$

and hence, by (6.21), the Green's function is $G(x,\tau - t;\xi)$, where

$$G(x,\tau - t;\xi) = 2\sum_{m=1}^{\infty}\mathrm{e}^{-m^2\pi^2(\tau - t)}\sin(m\pi x)\sin(m\pi\xi). \tag{6.38}$$

Of course, (6.38) can be written in a form similar to that of (6.36) by replacing every other hot spot by an equal and opposite 'cold spot'.[112] We also note that G can be written as a theta function (see Exercise 6.8 in the Neumann case).

We remarked in Chapter 5 that Fourier series representations give useful information about the behaviour of the solution as we move away from the boundary in elliptic problems. Here, the Fourier series expression for the Green's function is particularly useful for finding the long-time behaviour: as long as $\int_0^1 g(x)\sin(\pi x)\,\mathrm{d}x \neq 0$, for large values of t the dominant term in the series is that for the smallest m, namely $a_1\mathrm{e}^{-\pi^2 t}\sin\pi x$. This confirms our expectation that a body which starts hot cools exponentially in time if the temperature at its surface is maintained at zero. For an insulating boundary, in which $\partial u/\partial x = 0$ on $x = 0, 1$, the leading eigenvalue is zero and it is easy to show that $u \to \bar{g}$ as $t \to \infty$, where $\bar{g} = \int_0^1 g\,\mathrm{d}x$ is the average initial temperature (see Exercise 6.9). Both (6.37) and the solution of the Neumann problem indicate the smoothing effect of diffusion: the higher harmonics decay rapidly for large time, as presaged in Chapter 3. We note that series solutions such as (6.36) are more useful than (6.37) for estimating short-time behaviour, since then the 'physical' hot spot term dominates (except near the boundary, where there is also an image contribution); however, the converse is true for large times.

[112] The equivalence between the two is an example of the *Poisson summation formula*.

6.4.3.2 Boundary value problems in higher dimensions

Problems in higher dimensions are less likely to be susceptible to image methods, but can sometimes be approached using separation of the variables and eigenfunction expansions in the spatial variables. Consider, for example, the following Robin boundary value problem in a bounded domain:

$$\frac{\partial u}{\partial t} = \nabla^2 u \quad \text{in } \Omega, \qquad \frac{\partial u}{\partial n} + \alpha u = 0 \quad \text{on } \partial\Omega, \qquad u = g \quad \text{at } t = 0.$$

We have already made some general remarks about this problem following (6.16) but we can discover a lot more by writing the solution as

$$u(\boldsymbol{\xi}, \tau) = \sum_{m=0}^{\infty} g_m e^{-\lambda_m \tau} \phi_m(\boldsymbol{\xi}),$$

where the suitably normalised ϕ_m solve the Helmholtz eigenvalue problem

$$\nabla^2 \phi + \lambda\phi = 0 \quad \text{in } \Omega, \qquad \frac{\partial \phi}{\partial n} + \alpha\phi = 0 \quad \text{on } \partial\Omega,$$

with eigenvalues $\lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$, and $g_m = \int_\Omega \phi_m(\mathbf{x}) g(\mathbf{x})\, d\mathbf{x}$. Hence

$$u(\boldsymbol{\xi}, \tau) = \int_\Omega g(\mathbf{x}) \sum_{m=0}^{\infty} e^{-\lambda_m \tau} \phi_m(\mathbf{x})\phi_m(\boldsymbol{\xi})\, d\mathbf{x}. \tag{6.39}$$

This not only reveals the result that the Green's function is $G(\mathbf{x}, \tau - t; \boldsymbol{\xi})$, where

$$G(\mathbf{x}, t, \boldsymbol{\xi}) = \sum_{m=0}^{\infty} e^{-\lambda_m t} \phi_m(\mathbf{x})\phi_m(\boldsymbol{\xi}),$$

but it also predicts that $u(\mathbf{x}, t)$ tends to $g_0\phi_0(\mathbf{x})e^{-\lambda_0 t}$ as $t \to \infty$ provided $g_0 \neq 0$.

Now let us consider what this representation implies for the dependence of the solution on α. When α is positive, which corresponds to Newtonian cooling, the principal eigenvalue is positive, indicating that u decays to zero. However, with an 'energy input' boundary condition, possibly representing an active or controlled boundary, α is negative and so is the leading eigenvalue, so solutions grow exponentially in time; stability is lost but not well-posedness. We remark that, if $\partial\Omega = \partial\Omega_- \cup \partial\Omega_+$ with $\alpha \geqslant 0$ on $\partial\Omega_+$ and $\alpha \leqslant 0$ on $\partial\Omega_-$, then the relative sizes of the two parts of the boundary and the magnitudes of α determine the sign of λ_0, as illustrated by the following example.

Example 6.3 (*Robin data on a square*) Consider the solution of the heat equation in the square $-1 < x,\, y < 1$, with the boundary conditions

$$\frac{\partial u}{\partial n} + \alpha u = 0 \quad \text{on } y = \pm 1, \qquad \frac{\partial u}{\partial n} - \beta u = 0 \quad \text{on } x = \pm 1,$$

where $\alpha, \beta > 0$, so heat is put in on the sides $x = \pm 1$ and extracted from $y = \pm 1$.

The principal eigenfunction is $\phi_0 = \cosh \mu x \cos \nu y$, with

$$\lambda_0 = \nu^2 - \mu^2, \quad \mu = \tanh^{-1}\beta, \quad \nu = \tan^{-1}\alpha,$$

where $0 < \nu < \pi/2$. Thus $\lambda_0 = 0$ if $\tan^{-1}\alpha = \tanh^{-1}\beta$, while if $\beta > \tanh(\pi/2)$ then, no matter how large α is, $\lambda_0 < 0$ and the trivial solution is unstable.

6.4.3.3 Transform methods

When boundaries are absent in heat conduction problems, the spatial operator has a continuous spectrum, which is, of course, why we used Fourier transforms rather than Fourier series to obtain (6.30). Moreover, solutions are often sought over an infinite time interval $0 < t < \infty$, which also suggests the possibility of Laplace transforms. However, it can often be puzzling to try to decide the best variables in which to attempt a transform, and experience is the only guide to which transform is easiest technically. Hence we include some further examples of transform solutions.

Example 6.4 (*Fourier versus Laplace transforms* (i)) Suppose that, instead of taking a Fourier transform, we had tried to derive (6.32) by a Laplace transform in which

$$\tilde{u}(x,p) = \int_0^\infty u(x,t)e^{-pt}\,dt.$$

Then we would have obtained

$$\frac{d^2\tilde{u}}{dx^2} - p\tilde{u} = -g(x),$$

and, since we want u to decay as $|x| \to \infty$,

$$2\sqrt{p}\tilde{u}(x,p) = \int_{-\infty}^x g(\xi)e^{-\sqrt{p}(x-\xi)}\,d\xi + \int_x^\infty g(\xi)e^{\sqrt{p}(x-\xi)}\,d\xi,$$

where $\sqrt{p}$ is defined to have positive real part. In order to retrieve (6.32) we now have the irksome task of reversing the order of integration in the Laplace inversion of $\tilde{u}$, which involves using the fact that the inverse Laplace transform of $e^{-\sqrt{p}x}/\sqrt{p}$ is $e^{-x^2/4t}/\sqrt{\pi t}$.

Example 6.5 (*Fourier versus Laplace transforms* (ii)) Suppose we try to solve the boundary value problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \tag{6.40}$$

with

$$u(0,t) = h(t) \quad \text{for } t > 0 \qquad \text{and} \qquad u(x,0) = 0 \quad \text{for } x > 0,$$

by taking a Laplace transform in t. We find

$$\tilde{u}(x,p) = \tilde{h}(p)e^{-x\sqrt{p}},$$

where, as usual, $\Re\sqrt{p} > 0$. Hence, by the convolution theorem,

$$u(x,t) = \int_0^t h(s)w(x,t-s)\,\mathrm{d}s, \tag{6.41}$$

where w is such that $\tilde{w}(x,p) = \mathrm{e}^{-x\sqrt{p}}$. After some manipulations of the inversion integral, it can be shown that

$$w(x,t) = \frac{1}{\pi}\int_0^\infty \mathrm{e}^{-pt}\sin x\sqrt{p}\,\mathrm{d}p, \tag{6.42}$$

which can in turn be shown to be equal to $x\mathrm{e}^{-x^2/4t}/2\sqrt{\pi t^3}$ (see Exercise 6.13 for further discussion).

Things are much easier if, motivated by the nature of the spectrum of $\mathrm{d}^2/\mathrm{d}x^2$ with a zero Dirichlet condition at $x = 0$, we define the Fourier sine transform as in §5.7.2 by

$$\hat{u}_s(k,t) = \int_0^\infty u(x,t)\sin kx\,\mathrm{d}x.$$

This gives

$$\hat{u}_s(k,t) = \mathrm{e}^{-k^2t}\int_0^t kh(\tau)\mathrm{e}^{k^2\tau}\,\mathrm{d}\tau,$$

and, by inversion,

$$u(x,t) = \frac{2}{\pi}\int_0^\infty k\mathrm{e}^{-k^2t}\sin kx\left(\int_0^t h(\tau)\mathrm{e}^{k^2\tau}\,\mathrm{d}\tau\right)\mathrm{d}k,$$

which is just (6.41) combined with (6.42) when we reverse the order of integration and write $k = \sqrt{p}$.

Example 6.6 (*Another two-point boundary value problem*) Let us again solve the heat conduction equation, but now with the conditions

$$u(\pi,t) = g(t) \quad \text{for } t > 0, \qquad u(0,t) = 0 \quad \text{for } t > 0, \qquad u(x,0) = 0 \quad \text{for } 0 < x < \pi.$$

We could try separating the variables, but imposing the boundary condition at $x = \pi$ is difficult. We could use the Green's function (6.38), but the Laplace transform is even easier because

$$\tilde{u}(x,p) = \tilde{g}(p)\frac{\sinh\sqrt{p}x}{\sinh\sqrt{p}\pi},$$

and we have a straightforward convolution representation as long as we can invert $\sinh\sqrt{p}x/\sinh\sqrt{p}\pi$. The only singularities of this function are simple poles at $\sqrt{p} = \pm in$, $n = 1,2,3,\ldots$, and the corresponding residues of $\mathrm{e}^{pt}\sinh\sqrt{p}x/\sinh\sqrt{p}\pi$ are $(-1)^{n+1}(2n/\pi)\sin(nx)\mathrm{e}^{-n^2t}$. Hence

$$u(x,t) = \frac{2}{\pi}\sum_{n=1}^\infty \int_0^t (-1)^{n+1}n\sin(nx)\mathrm{e}^{-n^2s}g(t-s)\,\mathrm{d}s.$$

*6.4.4 Convection–diffusion problems

The inclusion of a convective term $(\mathbf{v} \cdot \nabla)u$ alongside the time derivative in the heat equation makes the spatial partial derivatives non-self-adjoint. This makes it awkward to use eigenfunction expansions because non-orthogonality means we have no obvious recipe for the coefficients. Of course, if there are no boundaries and $\mathbf{v}$ is constant, then it is possible to remove the convective term by changing position to $\mathbf{x} - \mathbf{v}t$. To illustrate the intricacies that can arise when $\mathbf{v}$ is not constant, we conclude this section by taking the Laplace transform of the simplified convection–diffusion equation (6.4) for pollutant flux in a river. For simplicity, the term $\partial^2 c/\partial x^2$ is neglected and the river is assumed to be two-dimensional, so that $c = c(x, y, t)$. We assume that the flow is in the x direction with speed $v(y) > 0$, that the river occupies $0 < y < 1$, $x > 0$, that there is no pollutant flux at the banks $y = 0, 1$ and that the water is initially pure. Thus the Laplace transform $\bar{c}(x, y, p) = \int_0^\infty c(x, y, t)\mathrm{e}^{-pt}\,\mathrm{d}t$ satisfies

$$D\frac{\partial^2 \bar{c}}{\partial y^2} = v(y)\frac{\partial \bar{c}}{\partial x} + p\bar{c} \quad \text{for } 0 < y < 1,\ x > 0,$$

with

$$\frac{\partial \bar{c}}{\partial y} = 0 \quad \text{at } y = 0, 1 \qquad \text{and} \qquad \bar{c} = \bar{c}_0(y, p) \quad \text{at } x = 0,$$

where $c_0(y, t)$ is the input of pollutant at $x = 0$.

This complicated problem for $\bar{c}$ can be best solved by an eigenfunction expansion resulting from separating the variables in x and y. The result is

$$\bar{c}(x, y, p) = \sum_{m=0}^{\infty} \bar{c}_{0m}\mathrm{e}^{-\lambda_m x}\psi_m(y, p), \tag{6.43}$$

where ψ_m are normalised eigenfunctions satisfying

$$D\frac{\mathrm{d}^2\psi_m}{\mathrm{d}y^2} - \left(p - \lambda_m v(y)\right)\psi_m = 0,$$

with

$$\frac{\mathrm{d}\psi_m}{\mathrm{d}y} = 0 \quad \text{at } y = 0, 1,$$

and

$$\bar{c}_{0m} = \int_0^1 \bar{c}_0\psi_m\,\mathrm{d}y.$$

It is only at all easy to find λ_m when p is small; this can be shown to correspond to t being large, which is often a limit that is of great practical interest. As shown in Exercise 6.19, for small p

$$\lambda_0 = \frac{p}{v_0} - \frac{D_0 p^2}{v_0^3} + \cdots,$$

where $v_0 = \int_0^1 v(y)\,\mathrm{d}y$, and D_0 is as given in Exercise 6.19. This means that the inverse Laplace transform of the first term in (6.43) satisfies the famous *Taylor diffusion* model

$$\frac{\partial c}{\partial t} + v_0 \frac{\partial c}{\partial x} = D_0 \frac{\partial^2 c}{\partial x^2};$$

we can see this simply by taking the Laplace transform of the Taylor diffusion model and examining its behaviour for small p. The effect of the y dependence of u, which represents the 'shear' in the flow, is thus equivalent to longitudinal diffusion in the x direction.

Convection–diffusion problems become even more difficult to solve when the velocity changes sign in the region of interest.[113] Suppose, for example, we consider the problem above in the steady state but with $v(y) = 2y - 1$. This means we have to solve a parabolic equation which is forwards in x in $y > \frac{1}{2}$ and backwards in x in $y < \frac{1}{2}$. Hence we do not expect to be able to prescribe an input everywhere at $x = 0$ but only the values of $c(0, y)$ for $\frac{1}{2} < y < 1$, and similarly $c(L, y)$ for $0 < y < \frac{1}{2}$ for some $L > 0$.[114] In practice, if we try to separate the variables as in the earlier examples, we find

$$c(x, y) = \sum_m A_m e^{-\lambda_m x} \psi_m(y),$$

where the eigenfunctions ψ_m satisfy

$$D\frac{\mathrm{d}^2\psi_m}{\mathrm{d}y^2} + \lambda_m(2y - 1)\psi_m = 0, \qquad \frac{\mathrm{d}\psi_m}{\mathrm{d}y}(0) = \frac{\mathrm{d}\psi_m}{\mathrm{d}y}(1) = 0.$$

In fact ψ_m is an *Airy function* (see Exercise 4.10) and the eigenvalues λ_m are both positive and negative, whereas they were all positive in the simpler cases discussed earlier. Although the ψ_m are complete in $0 < y < 1$, they are not complete in either $0 < y < \frac{1}{2}$ or $\frac{1}{2} < y < 1$, and hence the boundary conditions

$$c(0, y) = \sum_m A_m \psi_m(y), \qquad c(L, y) = \sum_m A_m e^{-\lambda_m L} \psi_m(y)$$

do not yield the coefficients A_m in a straightforward way.

It is an interesting attribute of parabolic equations that several of the examples above are susceptible to a totally different method of solution which can even be applied to nonlinear problems, and this is what we will describe in the next section.

[113] This happens frequently in industrial processes such as distillation, where 'counter-current' mass transfer takes place.

[114] There are many other more general 'forward-backward' problems, often called *Gevrey parabolic problems*, such as steady problems in which v changes sign across a general boundary.

6.5 Similarity solutions and groups

It is an immediate observation that, if $u(x,t)$ satisfies

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

then so does the function $u(\mu x, \mu^2 t)$, where μ is an arbitrary constant. This suggests the possibility of finding solutions for which

$$u(x,t) \equiv u(\mu x, \mu^2 t) \tag{6.44}$$

for all x and t. Now, for any particular value of t, we can always set $\mu = 1/\sqrt{t}$, so this identity means that $u = F(x/\sqrt{t})$ for some function F. In fact, when we write $\eta = x/\sqrt{t}$ we soon find

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\eta^2} + \frac{\eta}{2}\frac{\mathrm{d}F}{\mathrm{d}\eta} = 0, \quad \text{i.e.} \quad F = A\,\mathrm{erf}\left(\frac{\eta}{2}\right) + B,$$

where $\mathrm{erf}\, y$ is the *error function* $(2/\sqrt{\pi})\int_0^y \mathrm{e}^{-s^2}\,\mathrm{d}s$, and A and B are constants.

This observation puts us in a position to solve the heat equation subject to any initial/boundary conditions for which the data is invariant under the transformation

$$x = \frac{1}{\mu}x', \quad t = \frac{1}{\mu^2}t', \quad u = u'.$$

For example, with $u(0,t) = 1$ and $u(x,0) = 0$,

$$u(x,t) = 1 - \frac{2}{\sqrt{\pi}}\int_0^{x/2\sqrt{t}} \mathrm{e}^{-s^2}\,\mathrm{d}s = \mathrm{erfc}\left(\frac{x}{2\sqrt{t}}\right), \tag{6.45}$$

where $\mathrm{erfc}\, y$ is the *complementary error function.*

This solution procedure can be somewhat mystifying at first sight, so we now give a more systematic account of what are commonly called *similarity solutions.*

The basic idea is very simple. We just ask the following question.

> 'What changes of the dependent and/or independent variables make, or leave, the equation *autonomous* in one or more independent variable(s), i.e. cause those variable(s) to appear in the equation only through differentiation?'

The reason that this is interesting is that if the equation is autonomous in x then it has solutions independent of x. Of course, these are special solutions, but firstly they are much easier to find than the general solution, and secondly we may be lucky enough to be able to use them to fit the initial and boundary conditions. Indeed, we have seen this process in action in (4.15), where we reduced the number of independent variables from two to one and found the Riemann function of the telegraph equation by simply solving an ordinary differential equation. In another vein, even if our solution fails to satisfy all the requisite conditions, it might still give us an approximate answer in some limits, such as for small or large time.

Indeed, in (6.34) we have already seen the solution of a general initial value problem tending to a function that has a special form as time tends to infinity.

Although no systematic answer is available to the question just posed, one very helpful observation can be made. This comes from studying what happens to a function of just one variable, say $f(x)$, when we make an arbitrary change of variable. If $f(x)$ appears as part of the solution of a differential equation, and hence needs to be differentiated and subject to other manipulation, any change of variable $x' = g(x)$ in general generates all the complexities associated with 'functions of a function'. However, these mostly melt away if we consider a particular family of transformations,

$$x' = g(x, \lambda), \tag{6.46}$$

where λ is a continuously varying parameter[115] chosen so that $g(x, 0) = x$, provided that g satisfies one, admittedly stringent, condition. This is that

$$\frac{\partial g}{\partial \lambda}(x; \lambda) = F_1(\lambda) F_2(g(x; \lambda))$$

for some F_1 and F_2. If this is the case we can reparametrise so that

$$\frac{\partial g}{\partial \lambda} = F(g), \tag{6.47}$$

and, using the fact that $g(x, 0) = x$, we can see what happens to f. After a little manipulation of the relevant series, we find

$$\begin{aligned} f(x') &= f\left(g(x; \lambda)\right) \\ &= f\left(x + \lambda \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} + \frac{\lambda^2}{2!} \left.\frac{\partial^2 g}{\partial \lambda^2}\right|_{\lambda=0} + \cdots\right) \\ &= f(x) + \lambda \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} f'(x) \\ &\quad + \frac{\lambda^2}{2!}\left(\left.\frac{\partial^2 g}{\partial \lambda^2}\right|_{\lambda=0} f'(x) + \left.\left(\frac{\partial g}{\partial \lambda}\right)^2\right|_{\lambda=0} f''(x)\right) + \cdots . \end{aligned} \tag{6.48}$$

Now, if we define the *infinitesimal generator* $\mathcal{U}$ by the operation

$$\mathcal{U} f(x) = \left(\left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x) = F(x) f'(x), \tag{6.49}$$

then equation (6.48) collapses into

$$f(x') = f(x) + \lambda \mathcal{U} f(x) + \frac{\lambda^2}{2!} \mathcal{U}^2 f(x) + \cdots . \tag{6.50}$$

This is because

[115]The parameter μ in (6.44) is equal to e^{λ}, as we shall see below; this ensures that $g(x, 0) = x$.

$$\mathcal{U}^2 f(x) = \left(\left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} \frac{\mathrm{d}}{\mathrm{d}x} \right)^2 f(x)$$
$$= \left.\left(\frac{\partial g}{\partial \lambda}\right)^2\right|_{\lambda=0} f''(x) + \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} \frac{\mathrm{d}}{\mathrm{d}x} \left(\left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} \right) f'(x),$$

and we know that

$$\frac{\mathrm{d}}{\mathrm{d}x} \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} = F'(x)$$

and, from (6.47),

$$\left.\frac{\partial^2 g}{\partial \lambda^2}\right|_{\lambda=0} = F'(g) \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=0} = F'(x)F(x).$$

Thus (6.47) is the key ingredient that allows (6.48) to be written as (6.50); this procedure can be applied to all orders in λ and the series can be formally summed to

$$f(x') = \exp(\lambda\mathcal{U}) f(x).$$

To apply this idea to differential equations, we have to deal with at least two variables, whether dependent or independent, in which case (6.46) and (6.49) generalise to

$$x' = g(x, y; \lambda), \quad y' = h(x, y; \lambda), \tag{6.51}$$

with

$$\frac{\partial g}{\partial \lambda} = G(g, h), \quad \frac{\partial h}{\partial \lambda} = H(g, h), \tag{6.52}$$

and

$$g(x, y; 0) = x, \quad h(x, y; 0) = y, \tag{6.53}$$

$$\mathcal{U} = G|_{\lambda=0} \frac{\partial}{\partial x} + H|_{\lambda=0} \frac{\partial}{\partial y}. \tag{6.54}$$

6.5.1 Ordinary differential equations

Suppose that x and y are, respectively, the independent and dependent variables in an ordinary differential equation. The statement that the equation is autonomous in x is equivalent to saying that it is invariant under the transformation $x' = x+\lambda$, $y' = y$, which clearly satisfies (6.52). More generally, we might ask 'Could the given ordinary differential equation be made autonomous by a change of variables?' The answer is yes if it is invariant under a transformation (6.46), because we simply change to a new independent variable X and a new dependent variable Y such that $\mathcal{U}Y = 0$ and $\mathcal{U}X = 1$. This means that the differential equation for $Y(X)$ is invariant when X is translated by an arbitrary constant and Y is left undisturbed. For example, consider the linear equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + q(x)y = 0, \tag{6.55}$$

which is not autonomous but is invariant with $g = x$, $h = \mathrm{e}^{\lambda} y$, $G = 0$, $H = h$ and $\mathcal{U} = y\,\partial/\partial y$.[116] This tells us that, if we take, for example, $Y = x$ and $X = \log y$ in (6.55), then we obtain an equation autonomous in X; in fact, it is

$$\frac{\mathrm{d}^2 Y}{\mathrm{d}X^2}\Big/\left(\frac{\mathrm{d}Y}{\mathrm{d}X}\right)^3 - 1\Big/\left(\frac{\mathrm{d}Y}{\mathrm{d}X}\right)^2 = q(Y).$$

In this case we do not use the autonomy directly to seek a solution independent of X; rather we note that the autonomy ensures that the order of the differential equation can be lowered, say by considering $\mathrm{d}Y/\mathrm{d}X$ as a function of Y. Equally, suppose we happen to know that $y_0(x)$ satisfies (6.55). Then the equation is invariant with $g = x$ and $h = y + \lambda y_0(x)$. This tells us to take, say, $Y = x$ and $X = y/y_0(x)$, which is the well-known rule for lowering the order of a linear ordinary differential equation when one solution is available.

6.5.2 Partial differential equations

To keep things as simple as possible, we begin by just considering transformations of the independent variables. Thus, suppose that x and y are independent variables in a partial differential equation for $u(x, y)$ and we apply a transformation (6.51)–(6.53) which leaves the equation invariant. Now, in order to reduce the number of independent variables from two to one, all we have to do is solve for a new variable Y such that $\mathcal{U}Y = 0$, and transform to Y and any other independent variable, X say. We then have a partial differential equation which is invariant under a one-parameter change of variables in X; hence it has a solution in which $u = F(Y)$. For example, in the calculation before (4.15),

$$\frac{\partial^2 R}{\partial x \partial y} + R = 0,$$

the transformation $g = \mu x$, $h = y/\mu$ leaves the equation invariant. Then, writing $\mu = \mathrm{e}^{\lambda}$ gives

$$G = g = \mathrm{e}^{\lambda} x, \quad H = -h = -\mathrm{e}^{-\lambda} y, \quad \mathcal{U} = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y},$$

and hence $Y = xy$. More relevant for this chapter, the transformation $x' = G = g = \mathrm{e}^{\lambda} x$, $t' = H/2 = h = \mathrm{e}^{2\lambda} t$ leaves the equation $\partial u/\partial t = \partial^2 u/\partial x^2$ invariant, so

$$\mathcal{U} = x\frac{\partial}{\partial x} + 2t\frac{\partial}{\partial t};$$

thus, as observed at the beginning of this section, a possible choice of the 'autonomous' independent variable, which we called Y above, is x^2/t.

[116] We write $h = \mathrm{e}^{\lambda} y$ rather than $h = \lambda y$ for convenience, because it means that $\lambda = 0$ is the identity and no reparametrisation is needed to obtain (6.47).

Of course, similarity solutions can always be written down by inspection by people clever enough to spot them. The point about (6.51)–(6.53) is that it provides a relatively systematic procedure for finding the similarity solutions once g and h are available. This method is far from trivial, as will be illustrated on the following pages, and many computer packages are currently available for automating the procedure and some of its generalisations. Such generalisations will not be considered here except to remark that we could consider firstly 'extended' transformations in which the functions g and h depend upon the dependent variable and its derivatives as well as the independent variables, or, secondly, multiparameter transformations in which $g = g(x, \lambda_1, \lambda_2)$ and (6.47) is replaced by, say, $\partial g/\partial\lambda_1 = F^{(1)}(g)$ and $\partial g/\partial\lambda_2 = F^{(2)}(g)$.

Concerning terminology, the requirements (6.47) and (6.51)–(6.53), together with the existence of inverses, are conditions for g in (6.47) or (g, h) in (6.51)–(6.53) to form what is called a *Lie group* or *continuous group* of transformations. The trivial manipulations necessary to show that (6.47) is equivalent to the crucial group-closure condition, namely that, for all λ and μ, $g(g(x;\lambda);\mu) = g(x;\nu)$ for some $\nu = \nu(\lambda,\mu)$, are given in Exercise 6.20.

With these thoughts in mind, we are now in a position to describe a more systematic way of finding similarity solutions of the heat equation than that described at the beginning of this section. Suppose we consider a general continuous group of transformations of the independent variables

$$x' = f(x,t;\lambda), \quad t' = g(x,t;\lambda), \tag{6.56}$$

where the group parameter is such that $\lambda = 0$ is the identity, so that $x \equiv f(x,t;0)$ and $t \equiv g(x,t;0)$, and the closure operation is addition, which can be achieved without loss of generality as in (6.51)–(6.53). Thus, for small λ,

$$x' = x + \lambda U + \cdots, \quad t' = t + \lambda V + \cdots, \tag{6.57}$$

where the components U and V of the infinitesimal generator $\mathcal{U} = U\,\partial/\partial x + V\,\partial/\partial t$ are just functions of x and t. This means that, for small λ, the chain rules for changing the variables are, to $O(\lambda)$,

$$\begin{aligned}
\frac{\partial}{\partial t} &= \left(1 + \lambda\frac{\partial V}{\partial t}\right)\frac{\partial}{\partial t'} + \lambda\frac{\partial U}{\partial t}\frac{\partial}{\partial x'}, &\quad \frac{\partial}{\partial t'} &= \left(1 - \lambda\frac{\partial V}{\partial t}\right)\frac{\partial}{\partial t} - \lambda\frac{\partial U}{\partial t}\frac{\partial}{\partial x},\\
\frac{\partial}{\partial x} &= \lambda\frac{\partial V}{\partial x}\frac{\partial}{\partial t'} + \left(1 + \lambda\frac{\partial U}{\partial x}\right)\frac{\partial}{\partial x'}, &\quad \frac{\partial}{\partial x'} &= -\lambda\frac{\partial V}{\partial x}\frac{\partial}{\partial t} + \left(1 - \lambda\frac{\partial U}{\partial x}\right)\frac{\partial}{\partial x}.
\end{aligned} \tag{6.58}$$

We can now enforce the crucial invariance property by selecting U and V, and hence the group, by the condition that the heat conduction equation be left invariant under the transformation. Using the chain rule as above, the heat equation with x' and t' as independent variables is

$$\left(1+\lambda\frac{\partial V}{\partial t}\right)\frac{\partial u}{\partial t'}+\lambda\frac{\partial U}{\partial t}\frac{\partial u}{\partial x'}=\frac{\partial^2 u}{\partial x'^2}+\lambda\left(\frac{\partial^2 U}{\partial x^2}\frac{\partial u}{\partial x'}+\frac{\partial^2 V}{\partial x^2}\frac{\partial u}{\partial t'}\right) + 2\lambda\left(\frac{\partial U}{\partial x}\frac{\partial^2 u}{\partial x'^2}+\frac{\partial V}{\partial x}\frac{\partial^2 u}{\partial x'\partial t'}\right)+O(\lambda^2) \tag{6.59}$$

to lowest order[117] as $\lambda \to 0$, and so we require

$$\frac{\partial V}{\partial x}=0,\quad \frac{\partial U}{\partial t}=\frac{\partial^2 U}{\partial x^2},\quad \frac{\partial V}{\partial t}=\frac{\partial^2 V}{\partial x^2}+2\frac{\partial U}{\partial x}. \tag{6.60}$$

At first sight it looks as if we are no better off than we were with the heat conduction equation, but (6.60) is an overdetermined system of three equations in only two unknowns and it soon transpires that the only solutions are

$$U=\frac{cx}{2}+d,\quad V=ct+e, \tag{6.61}$$

where c, d and e are constants. Hence, when $d=e=0$, we can use $\partial x'/\partial\lambda = cx/2$ and $\partial t'/\partial\lambda = ct$ to find

$$x'=\mathrm{e}^{\lambda c/2}x,\quad t'=\mathrm{e}^{\lambda ct}t, \tag{6.62}$$

which gives (6.44) with a change in notation. Alternatively, with $c=0$, $d=v$ and $e=1$, we have

$$x'=x+\lambda v,\quad t'=t+\lambda, \tag{6.63}$$

leading to 'travelling wave' solutions of the form $u=F(x-vt)$; such a Galilean transformation is always possible for autonomous partial differential equations.

Concerning the procedure above, we note the following points.

1. The independent variables such as $x/\sqrt{t}$ and $x-vt$, corresponding to (6.62) and (6.63), respectively, could be enumerated quite systematically by solving for a variable ξ such that

$$\mathcal{U}\xi=\left(U\frac{\partial}{\partial x}+v\frac{\partial}{\partial t}\right)\xi=0. \tag{6.64}$$

2. In all cases u must satisfy a suitable ordinary differential equation. For (6.63), we find

$$\frac{\mathrm{d}^2F}{\mathrm{d}\xi^2}+v\frac{\mathrm{d}F}{\mathrm{d}\xi}=0,$$

so $F=A+B\mathrm{e}^{-v(x-vt)}$, where A and B are constants.

3. Systems such as (6.60) can easily look formidable, even though they are always linear and overdetermined. One of the boons of modern symbolic manipulations is that several packages are available to automate the process of deriving and solving such systems, and we will return to this in Chapter 9.

[117] In fact, using the group property (6.51)–(6.53) it can be shown that this implies invariance to all orders in λ.

4. It is relatively easy to generalise (6.57) to account for transformations of the dependent variable u. This is done in Exercise 6.21, where it is easily seen to lead to similarity solutions of the form

$$u = t^\alpha F\left(\frac{x}{\sqrt{t}}\right), \tag{6.65}$$

for any constant α. We recall that our fundamental 'heat source' solution (6.30) is of this form with $\alpha = -1/2$.

An important example is that of the fundamental solution of the heat equation in space dimension $m > 1$, which can be determined as one of these similarity solutions. Spherical symmetry demands that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{m-1}{r}\frac{\partial u}{\partial r},$$

where $r = |\mathbf{x}|$ and an obvious similarity solution is $u = t^\alpha F(r/\sqrt{t}) = t^\alpha F(\eta)$. The value of α must be such that the total amount of heat is conserved, i.e.

$$\int_{\mathbb{R}^m} u(\mathbf{x}, t)\,\mathrm{d}\mathbf{x} = t^\alpha \int_{\mathbb{R}^m} F\left(\frac{r}{\sqrt{t}}\right)\mathrm{d}\mathbf{x} = \text{constant} \cdot t^{\alpha+m/2}$$

is constant, and it is seen that $\alpha = -m/2$. The equation for F is now

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\eta^2} + \left(\frac{m-1}{\eta} + \frac{\eta}{2}\right)\frac{\mathrm{d}F}{\mathrm{d}\eta} + \frac{m}{2}F = 0. \tag{6.66}$$

Miraculously, there is an explicit solution $F = \text{constant} \cdot \mathrm{e}^{-\eta^2/4}$ which satisfies the regularity condition $\mathrm{d}F/\mathrm{d}\eta\,(0) = 0$. When we finally require $\int_{\mathbb{R}^m} u\,\mathrm{d}\mathbf{x} = 1$, we find that

$$u(\mathbf{x}, t) = (4\pi t)^{-m/2}\mathrm{e}^{-r^2/4t},$$

in accordance with (6.35).

Another generalisation is the inclusion of a constant source term in the heat equation, when again a similarity solution of the form (6.65) may be appropriate. For example, the half-line problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + 1 \quad \text{in } x > 0, \tag{6.67}$$

with $u = 0$ on $x = 0$ and $u = 0$ at $t = 0$, models uniform heating in a half-space with zero initial and boundary conditions. Now a similarity solution of the form (6.65) is vital to give us enough freedom to allow group invariance: (6.67) is invariant under the group $x' = \mathrm{e}^\lambda x$, $t' = \mathrm{e}^{2\lambda} t$, $u' = \mathrm{e}^{2\lambda} u$, and so we write $u = tF(\eta)$ and $\eta = x/\sqrt{t}$. This yields

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\eta^2} + \frac{\eta}{2}\frac{\mathrm{d}F}{\mathrm{d}\eta} - F + 1 = 0 \quad \text{for } \eta > 0,$$

with $F(0) = 0$ and, since $tF(x/t^{1/2}) \to 0$ as $t \to 0$, $F(\eta)/\eta^2 \to 0$ as $\eta \to \infty$. It can be shown that this is enough information for the solution to be written down in terms of parabolic cylinder functions (these are defined in Exercise 5.40).

*6.5.3 General remarks

Despite all the efforts that have been put into the development of a systematic theory for similarity solutions, several puzzling aspects remain. The principal shortcoming of the group invariance approach is that, at the end of the day, someone has to 'spot' the group under which the differential equation is invariant; the approach has made the identification task easier than that of spotting the similarity variables directly, but it has not automated it.

We wish to draw attention to two other aspects of the theory. The first is the relationship between the invariance approach and that of separation of the variables. The latter method has been freely used in the preceding chapters with little or no comment, and in every case of its use the reader has had to 'eyeball' the fact that separation was feasible. Group invariance sheds some light on this situation for, if we return to the argument leading to (6.65) and in particular to the calculation of Exercise 6.21, we soon see that the heat equation is invariant under the group whose infinitesimal generator is

$$\mathcal{U} = U\frac{\partial}{\partial x} + V\frac{\partial}{\partial t} + W\frac{\partial}{\partial u},$$

where one solution for U, V and W is

$$\begin{aligned} U &= \frac{1}{2}(A_1 + 2A_2t)x + A_3 + A_4t, \\ V &= A_5 + A_1t + A_2t^2, \\ W &= \left(A_6 - \frac{1}{2}A_2t - \frac{1}{2}A_4x - \frac{1}{4}A_2x^2\right)u, \end{aligned} \tag{6.68}$$

and A_i are constants. (The general solution has an arbitrary solution of the heat equation added to W.) Now it is easy to see that one possibility is

$$\mathcal{U} = A_3\frac{\partial}{\partial x} + A_6u\frac{\partial}{\partial u},$$

so $Y_1 = t$ and $Y_2 = ue^{-\alpha x}$, where $\alpha = A_6/A_3$, are both invariants for which $\mathcal{U}Y = 0$. Hence there is a solution in which $Y_2 = F(Y_1)$, i.e.

$$u = e^{\alpha x}F(t),$$

which is just the result of separation of variables!

Although we do not advocate this approach to separation of variables in this book, it is interesting to note that when we apply it to Laplace's equation, even with the restriction that only the independent variables x and y are candidates for transformation, we find that

$$x' = x + \lambda U + \cdots, \quad y' = y + \lambda V + \cdots,$$

where the infinitesimal generator $\mathcal{U} = U\,\partial/\partial x + V\,\partial/\partial y$, and that U and V satisfy the Cauchy–Riemann equations

$$\frac{\partial U}{\partial x} = \frac{\partial V}{\partial y}, \quad \frac{\partial V}{\partial x} = -\frac{\partial U}{\partial y}$$

(see Exercise 6.24). Hence, U and V are harmonic conjugates and we have recovered the conformal invariance of §5.9.1.

This leads to our final observation. Clearly, the more terms that enter a partial differential equation and its initial and boundary conditions, the less likely it is to admit a group invariance. Hence, similarity solutions are often of most practical value as representations of the asymptotic behaviour of the solution in the neighbourhood of certain interesting points in the space of independent variables at which the initial and boundary conditions take a simple form. For example, whenever we 'switch on' a boundary value for the solution of a parabolic equation that is different from its initial value, we expect the starting behaviour near this boundary to be described by the similarity solution (6.45), where x is the coordinate normal to the boundary. Such a result, if true, is of great value in ensuring that discretised versions of the problem start off in the right way, because no Taylor expansion exists near $t = 0$, $x = 0$. Equally, given any suitably localised initial distribution of heat on the line $-\infty < x < \infty$, we expect the temperature to tend to the similarity solution (6.34) as $t \to \infty$. This is because, at a large distance from the origin in the (x, t) plane, the initial data can be approximated by a multiple of the delta function.[118]

Such expectations are often correct but their proof is beyond the scope of this text. We will, however, mention one fascinating aspect of 'local similarity', which can best be illustrated by an example of an elliptic Dirichlet problem. Suppose

$$\nabla^2 u = 0$$

in two-dimensional polar coordinates, with

$$u = \begin{cases} 0, & \theta = 0,\ 0 < r < 1, \\ r^2, & \theta = \alpha,\ 0 < r < 1, \\ f(\theta), & r = 1,\ 0 < \theta < \alpha. \end{cases}$$

Assume, for the moment, that we are not clever enough to be able to solve the problem exactly by separation of the variables, yet we still seek to find the behaviour of u near the 'singular point' $r = 0$. A local similarity solution, obtained in practice by separation of variables, is $u \propto r^n \sin n\theta$, where, to satisfy the boundary conditions, we need $n = 2$ and

$$u = \frac{r^2 \sin 2\theta}{\sin 2\alpha}. \tag{6.69}$$

However, this approximation clearly fails at $\alpha = \pi/2$. All is revealed when we consider the exact solution

[118] If $f(x)$ is smooth and its integral is unity, then it is an easy exercise to show that $\lim_{\epsilon \to 0}(1/\epsilon) f(x/\epsilon) = \delta(x)$.

$$u = \frac{r^2 \sin 2\theta}{\sin 2\alpha} + \sum_{n=1}^{\infty} b_n r^{n\pi/\alpha} \sin\left(\frac{n\pi\theta}{\alpha}\right), \tag{6.70}$$

where the b_n are Fourier sine series coefficients defined by

$$f(\theta) - \frac{\sin 2\theta}{\sin 2\alpha} = \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi\theta}{\alpha}\right) \quad \text{for } 0 < \theta < \alpha.$$

The series (6.70) is valid for all α with $0 < \alpha < \pi$ as long as $\alpha \neq \pi/2$, and, indeed, can even be modified to work with $\alpha = \pi/2$ (see Exercise 6.25). Hence, it is clear that (6.69) is only the dominant term as $r \to 0$ when $\alpha < \pi/2$; for larger values of α, this term is swamped by the first Fourier term $b_1 r^{\pi/\alpha} \sin(\pi\theta/\alpha)$, which is still a similarity solution but one whose relevance is not obvious from a cursory glance at the differential equation and boundary conditions near $r = 0$. Indeed, its determination requires that we both solve an eigenvalue problem to find the requisite power of r and then use global information to find b_1. This is an example of what is called *second-kind similarity*. It is of great practical importance because it is often vital to know whether the local behaviour near a singularity of a solution of a partial differential equation is controlled by local rather than global considerations.

6.6 Nonlinear equations

6.6.1 Models

6.6.1.1 Semilinear equations

Reaction–diffusion equations are frequently semilinear and of the form

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u, x, t). \tag{6.71}$$

Typically they appear as models in population dynamics, with the inhomogeneous term depending upon a local population density, and in physical chemistry or chemical engineering, where f varies with temperature and/or chemical concentration in a reactor. The function f may be positive or negative to model say birth or death, or exothermic or endothermic reactions, respectively.

If the problem is homogeneous and autonomous, $f = f(u)$. For example, if u is the concentration of some chemical undergoing an Nth-order chemical reaction, $f(u) = \lambda u^N$. On the other hand, if u is the temperature in an exothermic reaction, f is often of the form $\lambda e^{u/(1+\epsilon u)}$. This is a rescaling[119] of the Arrhenius function introduced in §5.11.1 and, since ϵ is usually very small in practice, f is often replaced by λe^u, as in (5.159). Other autonomous models arise in situations where diffusion is trying to spread out the dependent variable uniformly in space while nonlinearity is trying to maintain it at a certain value. A famous example is the

[119]The Arrhenius function of temperature T is $e^{-E/RT}$, which can be written in the form $\lambda e^{u/(1+\epsilon u)}$ by setting $T = T_0(1 + \epsilon u)$, $\epsilon = RT_0/E$ and $\lambda = e^{-E/RT_0}$, where T_0 is a suitably chosen reference temperature.

Fisher equation for a population u in which $f = u - u^2$, the first term representing a linear birth rate and the second the limiting effect of the food supply, which diminishes as $u \to 1$. Another is the *Cahn–Allen equation* in which $f = u - u^3$; u is now the fraction of material that is undergoing a certain kind of phase change from one stable phase $u = -1$ to another stable phase $u = 1$.

Often semilinear equations contain small parameters multiplying f or the Laplacian in (6.71), and Exercise 6.26 indicates how such parameters can be exploited.

6.6.1.2 Quasilinear equations

Many problems in fluid dynamics lead to quasilinear rather than semilinear parabolic equations. A simple example concerns the flow of a compressible fluid through a porous medium. Taking u to represent the fluid density, then Darcy's law (5.26), which relates the velocity $\mathbf{v}$ to the pressure p by $\mathbf{v} = -(k/\mu)\nabla p$, the equation of conservation of mass, $\partial u/\partial t + \nabla \cdot (u\mathbf{v}) = 0$, and an equation of state, $p = p(u)$, can be combined to give

$$\frac{\partial u}{\partial t} = \nabla \cdot (K(u)\nabla u),$$

where $K(u)$ is proportional to $u\,\mathrm{d}p/\mathrm{d}u$. If we further specialise to the case of an isothermal perfect gas, we have $p \propto u$ while, for an adiabatic situation in which no heat is transferred to or from the gas, $p \propto u^\gamma$ with $\gamma > 1$. In either case the model can be written as the *porous medium equation*

$$\frac{\partial u}{\partial t} = \nabla \cdot (u^n \nabla u), \tag{6.72}$$

where n is a positive constant.

This equation also arises in the study of thermal radiation, where energy is transferred electromagnetically as well as by conduction or convection. Thus, at any point, there is not only the absolute temperature $T(\mathbf{x}, t)$, but also an electromagnetic energy density Q which depends on the direction $\mathbf{y}$ through the point $\mathbf{x}$, so we write

$$Q = Q(\mathbf{x}, \mathbf{y}, t), \quad |\mathbf{y}| = 1; \tag{6.73}$$

electromagnetic energy propagates at the speed of light c in all directions $\mathbf{y}$ in an isotropic material. In the absence of heat conduction, the key modelling assumptions are that the material emits electromagnetic energy at a rate βT^4 per unit volume,[120] where β is constant, and absorbs it at a rate proportional to Q. Thus, since c is large, the two energy balances are

$$c\mathbf{y} \cdot \nabla Q = \beta T^4 - \alpha Q, \tag{6.74}$$

where the gradient is with respect to $\mathbf{x}$, and

[120]The fourth power can be explained by the fact that the density of the energy emitted by a 'black body' as a function of the frequency ω of the radiation is, by quantum statistics, $8\pi h(\omega/c)^3/(e^{h\omega/kT} - 1)$, where h is Planck's constant, k is Boltzmann's constant and T is absolute temperature. Integrating over $0 < \omega < \infty$, we find that the total energy emitted is proportional to T^4.

$$\rho\bar{c}\frac{\partial T}{\partial t} = \int_{|\mathbf{y}|=1} (\alpha Q - \beta T^4)\,\mathrm{d}S = \alpha \int_{|\mathbf{y}|=1} Q\,\mathrm{d}S - 4\pi\beta T^4, \tag{6.75}$$

where ρ is the density, $\bar{c}$ is the specific heat and α is a constant. The presence of the integral in (6.75) means that, as is the case with most radiative transfer problems, the model is an integro-differential equation rather than a partial differential equation. However, it can be approximated by a differential equation in the so-called 'optically thick' limit in which β and α are large and comparable in an appropriate non-dimensionalisation. Then the dependence of Q on $\mathbf{y}$ is weak and, substituting iteratively in (6.74), we can write

$$Q = \frac{\beta}{\alpha}\left(T^4 - \frac{c}{\alpha}(\mathbf{y}\cdot\nabla)T^4 + \frac{c^2}{\alpha^2}(\mathbf{y}\cdot\nabla)^2 T^4 + \cdots\right). \tag{6.76}$$

Now

$$\int_{|\mathbf{y}|=1} T^4\,\mathrm{d}S = 4\pi T^4, \qquad \int_{|\mathbf{y}|=1} (\mathbf{y}\cdot\nabla)T^4\,\mathrm{d}S = \nabla(T^4)\cdot\int_{|\mathbf{y}|=1} \mathbf{y}\,\mathrm{d}S = 0$$

and

$$\int_{|\mathbf{y}|=1} (\mathbf{y}\cdot\nabla)^2 T^4\,\mathrm{d}S = \frac{4\pi}{3}\nabla^2(T^4),$$

so only the last term in (6.76) ultimately contributes to (6.75), which is simply (6.72) with $n = 3$ after a trivial rescaling.[121]

The porous medium equation also models the thickness h of a viscous drop spreading under gravity over a horizontal surface, as we can see by modifying the derivation of the paint model of §1.1. The horizontal velocity is again proportional to $y(2h - y)\nabla p$, but now p is nearly hydrostatic and thus approximately equal to $\rho g(h - y)$, so the horizontal components of ∇p and ∇h are, to leading order, proportional to each other. Finally, the equation of conservation of mass gives

$$\frac{\partial h}{\partial t} = \nabla(h^3\nabla h)$$

with a suitable scaling of time.

A similar equation can model the horizontal spreading of highly fissured volcanos, which can be considered as shallow porous media through which magma flows from below. The upper surface of the volcano moves normal to itself with a velocity proportional to the rate of arrival of the magma. Again taking h to be the height of the volcano surface above some horizontal datum, the magma pressure is also approximately hydrostatic, so that the Darcy velocity of the magma

[121] Unfortunately, radiative heat transfer in many processes such as glass manufacture is often midway between optically thick and optically thin, but porous medium equation models are still often used.

is proportional simply to $-\nabla h$. The total horizontal flow rate is then proportional to $-h\nabla h$ and conservation of mass means that

$$\frac{\partial h}{\partial t} = \nabla \cdot (h\nabla h),$$

and we can anticipate solutions of this equation in which h has the topography of, say, Mount Fuji.

Another example from fluid dynamics which leads to a quasilinear equation different from the porous medium equation is Burgers' equation (2.50), which arises in the theory of viscous one-dimensional gas dynamics; because of its importance, its detailed derivation is given in Exercise 6.27.

We conclude our modelling discussion by quickly mentioning another famous quasilinear equation which describes the two-dimensional steady flow of an incompressible fluid flowing in a boundary layer on a wall. This is the Prandtl equation for the stream function $\psi(x, y)$ which, for the case of flow past a flat plate, takes the form

$$\frac{\partial \psi}{\partial y}\frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial \psi}{\partial x}\frac{\partial^2 \psi}{\partial y^2} = \frac{\partial^3 \psi}{\partial y^3} \quad \text{for } 0 < y < \infty,\ x > 0, \tag{6.77}$$

with

$$\psi = \frac{\partial \psi}{\partial y} = 0 \quad \text{at } y = 0 \tag{6.78}$$

and

$$\psi = y + O(1) \quad \text{at infinity.} \tag{6.79}$$

In this model the fluid velocity is, as usual, $(\partial\psi/\partial y, -\partial\psi/\partial x)$ and some specification of the behaviour at the 'leading edge' $x = y = 0$ is necessary to close the problem. The derivation of (6.77)-(6.79) is too complicated to describe here but the basic idea will be given in Chapter 9. Suffice it to say that the left-hand side of (6.77) models the inertia of the flow, the right-hand side models viscous forces, (6.78) expresses the fact that the fluid adheres to the plate and (6.79) represents the uniform flow outside the boundary layer. While (6.77) is not strictly speaking parabolic, it can be made so by means of a partial hodograph transformation similar to that described in §4.8.2. When we regard $u = \partial\psi/\partial y$ as a function of x and ψ, the chain rule gives that

$$\frac{\partial^2 \psi}{\partial y^2} = u\frac{\partial u}{\partial \psi}, \quad \frac{\partial^2 \psi}{\partial x \partial y} = \frac{\partial u}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial u}{\partial \psi}, \quad \frac{\partial^3 \psi}{\partial y^3} = u\frac{\partial}{\partial \psi}\left(u\frac{\partial u}{\partial \psi}\right),$$

and hence (6.77) becomes

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial \psi}\left(u\frac{\partial u}{\partial \psi}\right),$$

which is just a special case of (6.72). However, in many applications it is easier to work with (6.77) directly.

6.6.2 Theoretical remarks

We have already remarked that almost all nonlinear partial differential equations, be they hyperbolic, elliptic or parabolic, need to be treated on their own merits. Concerning parabolic equations, a good rule is to start by asking the following questions.

1. Are there steady solutions?
2. Are there spatially homogeneous solutions?
3. Are there any symmetries that can be exploited to find similarity solutions, as in §6.5? (Questions 1 and 2 are really special cases of this.)

Other general questions which might also be posed are as follows.

4. Is the solution allowed to change sign? In all the examples above, except the Burgers and the Cahn–Allen equations, the physical interpretation demands that the dependent variable is non-negative.
5. Does the maximum principle apply? (When it does, it can be used to answer the previous question.)
6. Is there degeneracy, i.e. is the equation 'properly parabolic'? For example, the porous medium equation, which can be written as

$$\frac{\partial u}{\partial t} = u^n \nabla^2 u + n u^{n-1} |\nabla u|^2,$$

is not uniformly parabolic because as $u \to 0$ the coefficient of the elliptic operator $\nabla^2 u$ vanishes. Indeed, regarding the right-hand side as $u^{n-1}(u\nabla^2 u + n|\nabla u|^2)$, it is the gradient rather than the Laplacian of u that dominates when u becomes small, indicating that the equation looks more and more like a first-order equation in this limit.

For degenerate problems it is possible, as we shall see later in this section and again in Chapter 7, for there to be a *free boundary* separating the *support* of u, i.e. the region where $u > 0$, from the region where it vanishes. Of course, if such a free boundary does occur, say for the porous medium equation, we cannot expect the differential equation (6.72) to hold in its vicinity, because some derivatives of u almost certainly fail to exist. Hence, as in Chapter 1, we need to consider the possibility of defining a weak solution by integration in a suitably generalised sense.

The next two sections are devoted to some answers to these questions for equations of porous medium, reaction–diffusion and Burgers type.

6.6.3 Similarity solutions and travelling waves

Consider first the porous medium equation (6.72). It has no non-trivial spatially homogeneous solutions, and steady asymmetric solutions in more than one space dimension are hard to find, as we learned from the p-Laplacian problem of §5.11.1. However, in one space dimension we can find travelling waves which describe a free boundary separating a region in which $u > 0$, say $\xi = x - Vt < 0$, from one in which $u \equiv 0$. Such a phenomenon would have been quite impossible for

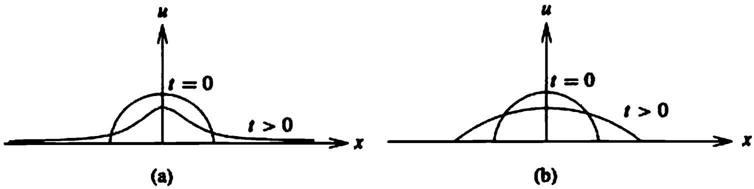


Fig. 6.3 Diffusion from an initially localised source: (a) non-degenerate; (b) degenerate.

a non-degenerate parabolic equation; for example, we know from (6.32) that the solution to a Cauchy problem for the heat equation is strictly positive for $t > 0$ for data that is positive only on a small interval and zero elsewhere (Fig. 6.3).

Writing $u = F(\xi)$ in $\partial u/\partial t = \partial/\partial x\,(u^n \partial u/\partial x)$ gives

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left(F^n \frac{\mathrm{d}F}{\mathrm{d}\xi}\right) + V\frac{\mathrm{d}F}{\mathrm{d}\xi} = 0,$$

so that one possibility is

$$u = \begin{cases} (nV(Vt - x))^{1/n}, & x < Vt, \\ 0, & x > Vt. \end{cases} \tag{6.80}$$

Note that, if $n \geqslant 1$, u ceases to be differentiable at $x = Vt$ and so the porous medium equation certainly does not hold there. However, we can see that

$$V = \lim_{x \uparrow Vt}\left(-\frac{1}{n}\frac{\partial}{\partial x}u^n\right), \tag{6.81}$$

which can be thought of as a Rankine–Hugoniot relation, as in §2.5. Indeed, this equation expresses conservation of mass at $x = Vt$ in the fluid dynamics interpretations of §6.6.1.2. In any case, the way is open to construct a theory of weak solutions of the porous medium equation by writing down the space–time integral of (6.72), after multiplication by suitable test functions, and seeking functions u that may be as badly behaved as (6.80), yet satisfy the integral identity. This is precisely the procedure that we adopted in Chapter 1, and, not surprisingly, we find that (6.81) is automatically satisfied by any weak solution. The analysis is too intricate to describe here but it is interesting that no 'entropy-like' criterion is now needed to ensure uniqueness.

Unexpected scenarios sometimes emerge when we seek travelling wave solutions of semilinear equations. For the autonomous case of (6.71), $\partial u/\partial t = \nabla^2 u + f(u)$, we encounter the ordinary differential equation

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2} + V\frac{\mathrm{d}F}{\mathrm{d}\xi} + f(F) = 0,$$

for which a phase-plane analysis can be carried out in terms of F and $\mathrm{d}F/\mathrm{d}\xi$. For the Fisher equation, where $f = F(1-F)$, Exercise 6.28 reveals that travelling waves are possible in which the population F tends to zero ahead of the wave ($\xi \to +\infty$), and to unity behind the wave ($\xi \to -\infty$), but that V must exceed[122] 2. The effect of the nonlinearity is to legislate against the slow waves in the continuous 'spectrum' $V \geqslant 0$ that exists for the heat equation with zero temperature as $\xi \to +\infty$.

A contrast emerges when we proceed to the semilinear equations with a cubic nonlinearity. When $f = u(u-a)(1-u)$ with $0 < a < 1/2$, so that $\int_0^1 f(u)\,\mathrm{d}u > 0$, the phase-plane in Exercise 6.29 reveals that there is a unique positive wave speed for which $u = F(\xi)$ connects the stable equilibrium points $F(\infty) = 0$ and $F(-\infty) = 1$. However, when $a = 1/2$, which is effectively the Cahn–Allen equation, $\int_0^1 f(u)\,\mathrm{d}u = 0$. Then the unique wave speed is also zero and the only travelling wave connecting $u = 0$ to $u = 1$ is the steady state.

We anticipated the structure of travelling wave solutions of Burgers' equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial x^2} \tag{6.82}$$

after (2.50). In Exercise 2.18 we effectively showed that, if $V > 0$ and

$$-V\frac{\mathrm{d}F}{\mathrm{d}\xi} + F\frac{\mathrm{d}F}{\mathrm{d}\xi} = \frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2} \tag{6.83}$$

and F takes any prescribed values $F_\pm$ as $\xi \to \pm\infty$, respectively, then

$$V = \frac{1}{2}(F_+ + F_-).$$

Hence the wave speed can take any value, but we also recall that the restriction $F_- > F_+$ must be imposed. Note the importance of the quadratic nonlinearity in (6.83). If a model ever led to Burgers' equation with a cubic nonlinearity, such as

$$\frac{\partial u}{\partial t} + u^2\frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial x^2},$$

we would find that travelling waves described by

$$-V\frac{\mathrm{d}F}{\mathrm{d}\xi} + F^2\frac{\mathrm{d}F}{\mathrm{d}\xi} = \frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2}$$

would not necessarily be constrained by the restriction $F_- > F_+$. This can be shown to allow the possibility of shock waves of expansion as well as compression, which would have serious implications for the discussion in §2.5.2.

The existence of all these travelling waves can be viewed, perhaps perversely, as resulting from the invariance of the relevant partial differential equations under arbitrary translations of x and t. Greater invariance is possible for equations with fewer terms, so let us now return to the porous medium equation (6.72). We just consider one-dimensional problems for simplicity (although many of the arguments

[122] It can be shown that the wave of minimum speed, 2, is the one that is observed as $t \to \infty$, because it has the best stability properties.

carry over to cases with radial symmetry) and we also restrict attention to $n = 1$ for ease of presentation. Then we can either follow the group theory analysis of §6.5 or simply observe that

$$u = t^{\alpha} F(\eta), \quad \eta = \frac{x}{t^{\beta}}$$

satisfies the porous medium equation as long as

$$t^{\alpha-1}\left(\alpha F - \beta\eta\frac{\mathrm{d}F}{\mathrm{d}\eta}\right) = t^{2\alpha-2\beta}\frac{\mathrm{d}}{\mathrm{d}\eta}\left(F\frac{\mathrm{d}F}{\mathrm{d}\eta}\right).$$

Hence $\beta = (\alpha+1)/2$ and

$$F\frac{\mathrm{d}^2F}{\mathrm{d}\eta^2} + \left(\frac{\mathrm{d}F}{\mathrm{d}\eta}\right)^2 + \beta\eta\frac{\mathrm{d}F}{\mathrm{d}\eta} - (2\beta-1)F = 0. \tag{6.84}$$

Group theory users have an advantage here because they can observe that (6.84) itself has an invariance which enables us to write

$$F = \eta^2 G, \quad \frac{\mathrm{d}F}{\mathrm{d}\eta} = \eta H,$$

and deduce the first-order equation

$$\frac{\mathrm{d}H}{\mathrm{d}G} = \frac{(G+\beta)H + (1-2\beta)G + H^2}{G(2G-H)}.$$

The phase-plane analysis of this equation allows us to describe a wide variety of solutions of the porous medium equation. For example, the critical point $G = -1/6$, $H = -1/3$, which corresponds to $F = -\eta^2/6$, enables us to construct the solution

$$u = \begin{cases} x^2/\left(6(t_0 - t)\right), & x \leqslant 0,\ t < t_0, \\ 0, & x \geqslant 0,\ t < t_0 \end{cases}$$

in $0 < t < t_0$, which is singular as $t \uparrow t_0$. Again, when $\beta = \frac{1}{3}$, $F = \frac{1}{6}(a^2 - \eta^2)$ gives

$$u = \begin{cases} \frac{1}{6}t^{-1/3}\left(a^2 - x^2/t^{2/3}\right), & |x| < at^{1/3}, \\ 0, & |x| > at^{1/3}, \end{cases} \tag{6.85}$$

which is even more interesting because it represents the spread of a 'blob' which was initially localised at $x = 0$, as in Fig. 6.3(b). In fact, this 'Barenblatt–Pattle' solution can be shown to tend to a multiple of $\delta(x)$ as $t \to 0$; it is also of the form (6.80) near the points $x = \pm at^{1/3}$.

There are other possibilities, including $\beta = 1$ which retrieves (6.80), but (6.85) has the most practical relevance. When written down in the radially-symmetric case, it can indeed predict the shape of Mount Fuji with good accuracy.

Perhaps the most dramatic, and certainly one of the most helpful, of all similarity reductions of parabolic equations occurs with the fearsome-looking Prandtl model (6.77). With hindsight it is quite easy to see its invariance under the transformation

$$x = e^{-2\lambda}x', \quad y = e^{-\lambda}y', \quad \psi = e^{-\lambda}\psi',$$

and hence the possibility of similarity solutions of the form

$$\psi = x^{1/2}F(\eta), \quad \eta = \frac{y}{x^{1/2}}.$$

This leads to the well-known boundary value problem for the *Blasius equation*

$$\frac{d^3F}{d\eta^3} + \frac{1}{2}F\frac{d^2F}{d\eta^2} = 0,$$

with $F = dF/d\eta = 0$ at $\eta = 0$ and $dF/d\eta \to 1$ as $\eta \to +\infty$. As with (6.84), the Blasius equation admits further group invariances, and it can in fact be written as a first-order equation by seeking $dF/d\eta$ as a function of F and then working with $\log F$ (see Exercise 6.31).

We conclude this section by making some brief remarks about the way in which similarity solutions and travelling waves may fit into the general structure of the solution of a Cauchy problem in which u is given at $t = 0$. Sometimes we encounter problems in which a small parameter distinguishes the magnitude of space and time derivatives. For example, we could consider the semilinear problem (6.71) in the form

$$\frac{\partial u}{\partial t} = \epsilon\nabla^2 u + f(u), \tag{6.86}$$

so that, when $\epsilon = 0$, the only steady states to which u could evolve are the stable zeros of $f(u)$. However, even in one space dimension, the perturbation analysis of (6.86) as $\epsilon \to 0$ is a difficult 'singular' perturbation problem. This is because, in some situations, the 'front' separating regions in which u is near different stable zeros of f moves exponentially slowly in ϵ.

In two space dimensions, the situation can become even more intriguing. Suppose we consider the Cahn–Allen equation in which $\int_0^1 f(u)\,du = 0$, and, for simplicity, restrict attention to problems with circular symmetry, so that

$$\frac{\partial u}{\partial t} = \epsilon\left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r}\right) + u - u^3.$$

Now suppose there is a front, which we denote by $r = R(t)$, separating the region $r > R$, in which u is very close to -1, from the region $r < R$, in which u is very close to 1. Near $r = R$ we expect u to change rapidly, as was the case when Burgers' equation was written as (2.50), so that we can approximate the solution by u_0, where

$$\frac{\partial u_0}{\partial t} = \epsilon\frac{\partial^2 u_0}{\partial r^2} + u_0 - u_0^3, \tag{6.87}$$

and u_0 is a travelling wave of the form

$$u_0 = F(r - R).$$

Now (6.87) is one equation for the two unknown functions f and R. However, as is common when parabolic equations are being approximated in thin regions, we can retrieve extra information from the correction to u_0, denoted by u_1. We find

$$-\dot{R}F'(r-R) + \frac{\partial u_1}{\partial t} = \epsilon\frac{\partial^2 u_1}{\partial r^2} + (1-3u_0^2)u_1 + \frac{1}{R}F'(r-R),$$

and it is intuitively clear[123] that u_1 can only be uniformly smaller than u_0 if

$$\dot{R} = -\frac{1}{R}.$$

This so-called *curvature flow* law, that the velocity of the front is proportional to its curvature, can even be shown to apply to non-circular configurations, and it gives a much simpler characterisation of the motion of the front than could ever be derived, say, for the motion of a shock wave in a hyperbolic problem. It is interesting to note that, if we write the front as $y = f(x,t)$ in a general situation, then the law becomes

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2} \Big/ \left(1 + \left(\frac{\partial f}{\partial x}\right)^2\right), \tag{6.88}$$

which is yet another quasilinear parabolic equation with fascinating properties.[124] We note that, if we solve this curvature flow equation when f is initially straight save for a small bump, then the straight segments move immediately, even though their initial velocity is zero everywhere, because diffusion takes place instantaneously. The curvature flow (6.88) also describes 'curve shortening'; rewriting it as in Exercise 6.33, it can be shown that the total length of the curve decreases with time.

Finally, let us return to the role of travelling waves in the solution of Burgers' equation. This gives us the opportunity to mention one of the most spectacular results about nonlinear parabolic equations, namely that the Cauchy problem for Burgers' equation in $-\infty < x < \infty$ can be solved exactly. We return to our group invariance discussion, where we noticed that the order of the ordinary differential equation $\mathrm{d}^2y/\mathrm{d}x^2 + y = 0$ could be lowered by changing to a new variable $\log y$. If we reverse the roles of the dependent and independent variables in that discussion so that $X = x$, this effectively results in the Ricatti equation $\mathrm{d}Z/\mathrm{d}x + Z^2 + 1 = 0$, where $Z = \mathrm{d}Y/\mathrm{d}x$. Hence, slightly changing the notation and working backwards, the equation $\mathrm{d}^2W/\mathrm{d}x^2 - W\,\mathrm{d}W/\mathrm{d}x = 0$ can be linearised by setting $W = -2\mathrm{d}/\mathrm{d}x\,(\log y)$. Let us therefore write[125]

$$u = -2\frac{\partial}{\partial x}\log v$$

[123] This very shaky argument about the particular integral for u_1 can be systematised by using the Fredholm Alternative.

[124] If the front velocity is proportional to the negative curvature, we obtain a realisation of a nonlinear backwards heat equation.

[125] This transformation is the same as the one we used on Liouville's equation in Chapters 4 and 5.

in Burgers' equation (6.82), the famous *Hopf-Cole transformation*. We obtain

$$\frac{\partial}{\partial x}\left(\frac{1}{v}\left(\frac{\partial v}{\partial t}-\frac{\partial^2 v}{\partial x^2}\right)\right)=0,$$

and so, if v satisfies the heat equation with initial condition

$$v(x,0)=\exp\left(-\frac{1}{2}\int^x g(\xi)\,\mathrm{d}\xi\right),$$

then $u=-(2/v)\,\partial v/\partial x$ is the solution of Burgers' equation with $u(x,0)=g(x)$. Thus, steady solutions of Burgers' equation can be identified with separable solutions of the heat equation.

6.6.4 Comparison methods and the maximum principle

One tool that is very useful for the qualitative study of certain kinds of semilinear parabolic equations is the comparison method, which is based on the maximum principle. The idea is a generalisation of that introduced in §5.11.2.1 for nonlinear elliptic problems. Taking, for example, the semilinear problem

$$\begin{aligned}\frac{\partial u}{\partial t}&=\nabla^2 u+f(u)\quad \text{in } \Omega,\\ \frac{\partial u}{\partial n}+\alpha u=0\quad \text{on } \partial\Omega,&\qquad u=g\quad \text{at } t=0,\end{aligned}\tag{6.89}$$

then $\underline{u}$ is said to be a lower solution if

$$\frac{\partial \underline{u}}{\partial t}\leqslant \nabla^2\underline{u}+f(\underline{u})\quad \text{in } \Omega,\qquad \frac{\partial \underline{u}}{\partial n}+\alpha\underline{u}\leqslant 0\quad \text{on } \partial\Omega,\qquad \underline{u}\leqslant g\quad \text{at } t=0.\tag{6.90}$$

Similarly, if $\overline{u}$ satisfies the reverse inequalities, then $\overline{u}$ is an upper solution. As long as f is Lipschitz continuous, the strong maximum principle that we described in §6.3 can be applied directly to $u-\underline{u}$ and $\overline{u}-u$ to guarantee that $\underline{u}\leqslant u\leqslant \overline{u}$. This means that local existence, uniqueness and continuous dependence on the data can sometimes be determined via a monotone iteration scheme starting from either $\underline{u}$ or $\overline{u}$ and based on the Picard theorem for ordinary differential equations. As in §5.11.2.1, we have to assume that a constant K can be chosen so that $f(u)+Ku$ is increasing; then we can consider the iteration given by

$$\begin{aligned}\frac{\partial u_n}{\partial t}-\nabla^2 u_n+Ku_n&=f(u_{n-1})+Ku_{n-1}\quad \text{in } \Omega,\\ \frac{\partial u_n}{\partial n}+\alpha u_n=0\quad \text{on } \partial\Omega,&\qquad u_n=g\quad \text{at } t=0,\end{aligned}$$

for $n\geqslant 1$. The starting point can be taken as either $u_0=\underline{u}$ or $u_0=\overline{u}$. Now, since

$$\frac{\partial}{\partial t}(u_1-u_0)-\nabla^2(u_1-u_0)+K(u_1-u_0)\geqslant 0,$$

on taking $u_0=\underline{u}$, the argument after (6.90) applied to u_1-u_0 shows that $u_1\geqslant u_0$. The strong maximum principle can then be applied inductively to show that, if

$u_{n-1} \leqslant u_n \leqslant \overline{u}$, then $u_n \leqslant u_{n+1} \leqslant \overline{u}$. It follows that u_n is an increasing sequence bounded by $\overline{u}$, and hence that $u_n \to u$ as $n \to \infty$ for some u between $\underline{u}$ and $\overline{u}$. The proofs are generalisations of the arguments in §5.11.2. The fact that u is a solution to the initial-boundary value problem can be deduced, assuming the existence of a Green's function G, by constructing an iteration scheme for an integral equation of the form

$$u(\mathbf{x},t) = \int_\Omega Gg\, d\boldsymbol{\xi} + \int_0^t \int_\Omega G\left(f(u) + Ku\right) d\boldsymbol{\xi}\, d\tau; \tag{6.91}$$

the uniqueness of the solution u follows as with the Picard-type argument for ordinary differential equations.[126]

Example 6.7 (*The Fisher equation*) Consider the following problem:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u(1-u) \quad \text{for } 0 < x < 1, \tag{6.92}$$

with $u = 0$ on $x = 0$ and $x = 1$, and $u = \sin \pi x$ at $t = 0$.

Obvious upper and lower solutions are $\overline{u} = 1$ and $\underline{u} = 0$, respectively, but a better choice of upper solution is $\overline{u} = Ae^{-\alpha t} \sin \pi x$, with $A \geqslant 1$ and $0 < \alpha \leqslant \pi^2 - 1$. The latter choice immediately shows that $u \to 0$, and the population tends to extinction, as $t \to \infty$.

This example suggests that, in more general semilinear cases, bounds on the solution can be determined simply by using the leading eigenfunction for the problem that we obtain when we linearise about a steady state. Thus, suppose that U is a steady state of

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u) \quad \text{in } \Omega, \qquad u = h(x) \quad \text{on } \partial\Omega, \tag{6.93}$$

and we linearise by writing $u = U + \tilde{u}$ to give, approximately,

$$\frac{\partial \tilde{u}}{\partial t} = \nabla^2 \tilde{u} + f'(U)\tilde{u} \quad \text{in } \Omega, \qquad \tilde{u} = 0 \quad \text{on } \partial\Omega. \tag{6.94}$$

In order to use the familiar technique of linear stability theory, to which we have already alluded in §5.11.3, we would now have to determine the real principal eigenvalue μ_0 of the spectral problem

$$\nabla^2 \phi + f'(U)\phi + \mu\phi = 0 \quad \text{in } \Omega, \qquad \phi = 0 \quad \text{on } \Omega; \tag{6.95}$$

if μ_0 is positive we expect the steady state to be stable. However, in this case, we can prove the stability by our comparison method because we can exploit the

[126] As usual, once we have a representation like (6.91) for the solution of any partial differential equation, we have effectively reduced the problem to one for an ordinary differential equation, as will be discussed further in Chapter 9.

single-signedness of the principal eigenfunction ϕ_0, shown in §5.7.1. Taking $\phi_0 \geqslant 0$, we pick a number c to be so small that

$$f(U - ce^{-\beta t}\phi_0) - f(U) - ce^{-\beta t}\phi_0 f'(U)$$

is of $o(c)$ as $c \to 0$ for some β with $0 < \beta < \mu_0$. Then, if we set

$$\underline{u} = U - ce^{-\beta t}\phi_0, \quad \overline{u} = U + ce^{-\beta t}\phi_0,$$

we find that

$$\frac{\partial \underline{u}}{\partial t} - \nabla^2 \underline{u} - f(\underline{u}) = c(\beta - \mu_0)e^{-\beta t}\phi_0 + o(c),$$
$$\frac{\partial \overline{u}}{\partial t} - \nabla^2 \overline{u} - f(\overline{u}) = c(\mu_0 - \beta)e^{-\beta t}\phi_0 + o(c),$$

so that, if c is positive, $\underline{u}$ and $\overline{u}$ are lower and upper solutions, respectively. Finally, taking initial data with $U - c\phi_0 \leqslant u(\mathbf{x}, 0) \leqslant U + c\phi_0$, we deduce that

$$U - ce^{-\beta t}\phi_0 \leqslant u \leqslant U + ce^{-\beta t}\phi_0,$$

and hence that $u \to U$ as $t \to \infty$.

We can make one general remark about the case when $h = h_0 =$ constant in (6.93) with $U \equiv h_0$. Then (6.95) is just Helmholtz' equation, which has positive eigenvalues. Thus, as long as $f'(h_0) \leqslant 0$, we can be sure that $\mu_0 > 0$, and so diffusion can never destabilise a stable solution of the spatially homogeneous problem. In §6.7.2, this will be seen *not* to be true for a parabolic system.

Precisely the same ideas can be used to prove instability for problems for which the linearisation has a negative principal eigenvalue, but, in cases where $\mu_0 = 0$, the linearised problem gives the least information about stability. Then the technique of upper and lower solutions is even more important, as shown by the example

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u^3 \quad \text{for } 0 < x < 1, \tag{6.96}$$

with $\partial u/\partial x = 0$ on $x = 0$ and $x = 1$. Solutions to the spatially homogeneous problem are now exact solutions to the parabolic equation and boundary conditions, and hence they serve as upper and lower solutions for initial–boundary value problems. Thus, the equilibrium $u \equiv 0$ is unstable.

The discussion above shows that there is plenty of scope for using comparison theorems and information about the eigenvalues of Helmholtz' equation to infer the stability or otherwise of steady states of semilinear parabolic equations. Another example is given in Exercise 6.35, but here we will just mention a specific situation that has important implications for exothermic combustion theory, as modelled after (6.71). When $\lambda f(u) > 0$, the 'smallest' steady solution U of

$$\frac{\partial u}{\partial t} = \nabla^2 u + \lambda f(u) \quad \text{in } \Omega \qquad \text{with} \qquad u = 0 \quad \text{on } \partial\Omega$$

is stable. The proof goes as follows.

From our discussion in §5.11.3, we assume that U depends continuously on λ in some interval $0 < \lambda < \Lambda$. Hence we take λ_i such that $0 < \lambda_1 < \lambda_2 < \Lambda$ and let the corresponding steady states U_i serve as lower and upper solutions for the evolution problem with $\lambda = \lambda_0$, $\lambda_1 < \lambda_0 < \lambda_2$. Then, if the initial condition is sufficiently close to U_0, so that

$$U_1 \leqslant u(\mathbf{x}, 0) \leqslant U_2,$$

then the comparison method ensures that u remains between U_1 and U_2. Hence, the stability of the steady state U_0 follows when we let $\lambda_1 \uparrow \lambda_0$ and $\lambda_2 \downarrow \lambda_0$.

*6.6.5 Blow-up

We have frequently seen in this book that nonlinearity can generate singularities in the solutions of well-posed partial differential equations, be they elliptic, hyperbolic or parabolic. Such singularity development is a global phenomenon, depending on data quite remote from the singularity location, and this aspect makes prediction difficult.

For semilinear equations, some very helpful clues can be found by studying the monotonicity of the solution in time using the comparison method. Suppose, for example, that in the Robin problem (6.89) the initial data $u(\mathbf{x}, 0) = g(\mathbf{x})$ is a lower solution, i.e. that it satisfies (6.90) with $\partial \underline{u}/\partial t = 0$. Hence $u(\mathbf{x}, t) \geqslant g(\mathbf{x})$ and we can show[127] that $\partial u/\partial t \geqslant 0$ for a short time. Thus we can take $u(\mathbf{x}, h)$ as a new subsolution starting from a small enough time $t = h$. Repeating the process shows that $\partial u/\partial t$ stays positive and hence, if we can additionally find a smallest steady state greater than g, then u must tend to this steady state as $t \to \infty$.

These arguments reveal the following general alternative for semilinear scalar problems: if g is a lower solution, then, because u is increasing, either u tends to the smallest equilibrium above g, or u is unbounded. Consider, for example, the ignition model of §6.6.1 in which

$$\frac{\partial u}{\partial t} = \nabla^2 u + \lambda e^u \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \qquad u = 0 \quad \text{at } t = 0. \tag{6.97}$$

We recall from §5.11.3.1 that the steady-state behaviour is characterised by the existence of a λ^* such that the continuous minimal branch w of positive solutions emanating from the origin in the response diagram first turns over at $\lambda = \lambda^*$. For the evolution problem, the initial condition zero is certainly a strict lower solution, so $\partial u/\partial t > 0$ for $\lambda > 0$. Hence, if $\lambda < \lambda^*$, then $u \to w$ as $t \to \infty$. In fact, we can extend the argument above to show that the smallest equilibrium state is stable both from above and below.

In contrast, if $\lambda > \lambda^*$ in (6.97), and indeed if $\lambda \leqslant \lambda^*$ and $u(\mathbf{x}, 0)$ is too large, there is the possibility that u is unbounded as $t \to \infty$. Moreover, it may well happen that u goes to infinity in finite time, in which case we say that *blow-up* has occurred. To demonstrate the inevitability of blow-up when λ is sufficiently

[127]To do this, we need to note that $u(\mathbf{x}, h) \geqslant u(\mathbf{x}, 0)$ for $h \geqslant 0$. By considering the evolution of $u_h(\mathbf{x}, t) = u(\mathbf{x}, t + h)$, we can see that $u_h(\mathbf{x}, t) \geqslant u(\mathbf{x}, t)$ and hence, taking the limit $h \downarrow 0$, we find that $\partial u/\partial t \geqslant 0$.

large, we need only consider the behaviour of the leading Fourier coefficient of the expansion of u in terms of the eigenfunctions of the Helmholtz problem

$$\nabla^2\phi_i + \mu_i\phi_i = 0, \qquad \phi_i = 0 \quad \text{on } \partial\Omega. \tag{6.98}$$

Taking ϕ_0 to be positive and such that $\int_\Omega \phi_0 \, \mathrm{dx} = 1$ without loss of generality, and setting $a = \int_\Omega u\phi_0 \, \mathrm{dx}$, we find

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \int_\Omega \phi_0 \frac{\partial u}{\partial t} \, \mathrm{dx} = \int_\Omega \phi_0(\lambda \mathrm{e}^u + \nabla^2 u) \, \mathrm{dx} = \int_\Omega \phi_0(\lambda \mathrm{e}^u - \mu_0 u) \, \mathrm{dx},$$

by Green's theorem. Now, noting that, by Jensen's inequality (see footnote 97 on p. 216)

$$\int_\Omega \phi_0 \mathrm{e}^u \, \mathrm{dx} \geqslant \exp\left(\int_\Omega \phi_0 u \, \mathrm{dx}\right),$$

we have

$$\frac{\mathrm{d}a}{\mathrm{d}t} \geqslant \lambda \mathrm{e}^a - \mu a,$$

which clearly implies that a blows up if λ is sufficiently large.

Such arguments can be used even when $\lambda < \lambda^*$ to show that blow-up also occurs if the initial data g is sufficiently large (see Exercise 6.35 and (6.99) below). In fact, many further aspects of blow-up can be addressed, such as the behaviour near $\lambda = \lambda^*$ or the prediction of the spatial variation of u near blow-up. The latter is of especial practical importance because of the occurrence of hot spots in large stores of solid material which may be undergoing even a gentle exothermic reaction.[128]

Rather than go into these intricate details here, we conclude by briefly mentioning one other qualitative approach to blow-up, namely the use of integral estimates. These may require considerable ingenuity but, for the semilinear equation $\partial u/\partial t = \nabla^2 u + f(u)$, with zero Dirichlet data, we can multiply by $\partial u/\partial t$ and integrate to give

$$\begin{aligned} 0 \leqslant \int_\Omega \left(\frac{\partial u}{\partial t}\right)^2 \mathrm{dx} &= \int_\Omega \left(f(u)\frac{\partial u}{\partial t} - \nabla u \cdot \nabla \frac{\partial u}{\partial t}\right) \mathrm{dx} \\ &= -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \left(|\nabla u|^2 - F(u)\right) \mathrm{dx}, \end{aligned}$$

where $F(u) = 2\int^u f(s)\,\mathrm{d}s$. Thus the 'energy'

$$E = \int_\Omega \left(|\nabla u|^2 - F(u)\right) \mathrm{dx}$$

is decreasing. We now use the Rayleigh–Ritz characterisation (see §5.7.1) of the positive principal eigenvalue μ_0 in (6.98) as $\min \int_\Omega |\nabla u|^2 \, \mathrm{dx} \big/ \int_\Omega u^2 \, \mathrm{dx}$. This implies that $E \geqslant \int_\Omega \left(\mu_0 u^2 - F(u)\right) \mathrm{dx}$ and all we need is for this integral to be

[128] Such hot spots have even been proposed as a mechanism for spontaneous human combustion.

positive for small u to guarantee stability of the zero solution. This is the case if, say, $f(u) = u^p$, $p > 1$.

On the other hand, suppose we consider

$$J = \frac{1}{2}\int_\Omega u^2\,\mathrm{dx}.$$

Now, a simple calculation shows that, when $f(u) = u^p$, again with $p > 1$,

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \int_\Omega \left(u^{p+1} - |\nabla u|^2\right)\mathrm{dx} = \frac{p-1}{p+1}\int_\Omega u^{p+1}\,\mathrm{dx} - E(t).$$

We can estimate $\int_\Omega u^{p+1}\,\mathrm{dx}$ in terms of J by again using Jensen's inequality and the monotonicity of E to give

$$\frac{\mathrm{d}J}{\mathrm{d}t} \geqslant \text{constant} \cdot J^{(p+1)/2} - E(0). \tag{6.99}$$

Hence, if $u(\mathbf{x}, 0)$ is large enough that $E(0) < 0$, we again have finite-time blow-up.

As a postscript to this section, we remark that blow-up can occur even for *linear* parabolic problems, as was hinted at in the discussion after (6.33). Our remarks there show that blow-up is generic for the backwards heat equation, but that it can also occur for forward equations is revealed by setting[129]

$$u = v\mathrm{e}^{x^2/2} \tag{6.100}$$

in the heat equation $\partial u/\partial t = \partial^2 u/\partial x^2$. This yields the seemingly innocuous forward equation

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + 2x\frac{\partial v}{\partial x} + (x^2+1)v. \tag{6.101}$$

Hence, if we seek solutions in which $v = g(x)$ at $t = 0$, we obtain from (6.30) and (6.100) that

$$v(x,t) = \frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} g(\xi)\mathrm{e}^{(\xi^2-x^2)/2-(\xi-x)^2/4t}\,\mathrm{d}\xi$$

instead of (6.32), which, assuming that $g(x)$ does not decay too fast as $|x| \to \infty$, clearly blows up when the sign of $1/2 - 1/4t$ changes at $t = 1/2$. This behaviour, which may seem unexpected from a casual glance at (6.101), can be interpreted either in terms of the fact that the solution $\mathrm{e}^{x^2/4(1-t)}/\sqrt{1-t}$ of the heat equation blows up as $t \uparrow 1$ because of its growth at infinity, or because of the unbounded spatial variation of the coefficients in (6.101).

*6.7 Higher-order equations and systems

We conclude this arduous but important chapter by citing some more exotic parabolic equations which further emphasise the sensitive dependence of the solutions on the data. Our illustrations involve models of physical processes that

[129]The authors are grateful to Dr R. Hunt for this remark.

lead to equations of order higher than two, either in scalar or vector form. However, we will restrict attention to equations of first order in time and even order in space because they are the only ones which bear passing resemblance to those discussed earlier in the chapter. Scalar third-order equations, for example, can exhibit oscillatory behaviour that is in many ways more reminiscent of the solution of hyperbolic problems, as we shall see in Chapter 9.

Even with our self-imposed restrictions, we still encounter a problem of presentation that has already become apparent in the previous chapter. This is the fact that the higher order a class of differential equations becomes, the less easy it is to make any general statements about the properties of the solutions. Thus, particular examples of the class must increasingly be studied individually. In order for this section not to degenerate into ever more detailed accounts of specific problems, we will only give the reader a glimpse of the possible pitfalls and what is possible concerning modelling and methodology. Hence we will do scant justice to what can often be achieved in any special case.

6.7.1 Higher-order scalar problems

One source of fourth-order parabolic equations is via the 'regularisation' of ill-posed second-order problems. For example, as mentioned in §6.6.1, it is popular to model phase separation in solids by introducing u to represent the fraction of material that has transformed but, rather than use the Cahn–Allen model, to assume that

$$\frac{\partial u}{\partial t} = \nabla^2 f(u), \tag{6.102}$$

where f is related to u using statistical thermodynamics. In fact, $f(u)$ is the derivative of the so-called *free energy* $F(u)$ of the material, as shown in Fig. 6.4, the 'potential wells' at $u = \pm 1$ representing the distinct phases; the regions in which $|u| > 1$ are unphysical.

If the boundary and initial data lie entirely within one phase, i.e. near $u = -1$ or near $u = 1$, then the coefficient of $\nabla^2 u$ in the right-hand side of (6.102), which

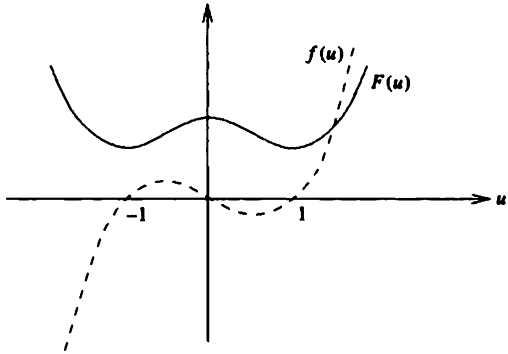


Fig. 6.4 Free energy for phase separation.

is simply $f'(u)$, remains positive, the maximum principle applies, and the problem is well posed. But then there is no change of phase. To get different phases the solution must contain regions where u is near both ± 1, and hence, in places, u takes values for which $f' < 0$, the so-called 'spinodal' region. In this region, the equation is a backward heat equation, and if the initial data is also spinodal the problem is ill-posed. We comment that, for times before the blow-up that would generally occur in the spinodal region, maxima tend to increase and minima to decrease. Thus u changes so that it abhors the spinodal region and takes values that increasingly correspond to phase separation.

In any event, (6.102) needs to be regularised, and one way in which this can be done is by including a higher-order term with the correct sign, and one such regularised equation is

$$\frac{\partial u}{\partial t} = \nabla^2 f(u) - \epsilon^2 \nabla^4 u,$$

where ϵ is a small constant. This fourth-order parabolic equation is called the *Cahn–Hilliard equation*, and the final biharmonic term can be identified with the energy density contained in the surface of the interface between the phases. The reason for the choice of sign, and, indeed, a guide to the well-posedness of such an equation, is given by considering a linear one-dimensional problem with oscillatory initial data in the spirit of §3.1. Consider

$$\frac{\partial u}{\partial t} = -\frac{\partial^2 u}{\partial x^2} - \epsilon^2 \frac{\partial^4 u}{\partial x^4} \quad \text{for } t > 0 \qquad \text{with} \qquad u(x,0) = \sin nx. \tag{6.103}$$

This has the solution $u(x,t) = e^{\lambda t} \sin nx$, with $\lambda = n^2(1 - \epsilon^2 n^2)$. Although low wavenumbers in which $n < 1/\epsilon$ give rise to temporal growth, solutions with high wavenumbers with $n > 1/\epsilon$ decay; the zero solution is unstable, which it was not in the case of the heat equation, but there is no longer any indication of arbitrarily rapid growth, as happens with the backward heat equation. Likewise, as long as ϵ is real, the term $-\epsilon^2 \nabla^4 u$ in the Cahn–Hilliard equation acts to limit rapid variations of u, confirming the expectation that the material reduces its interfacial energy. The lack of stability to low wavenumber perturbations is a clear indication that there is no maximum principle for this particular equation. Although instability and the failure of the maximum principle for (6.103) result from the $-\partial^2 u/\partial x^2$ term, the maximum principle in fact also fails for $\partial u/\partial t = -\partial^4 u/\partial x^4$, as we can see by taking

$$u = 105 - 24t > 0 \quad \text{for } x = \pm 3,\ 0 \leqslant t \leqslant 2$$

and

$$u = x^4 + 24 > 0 \qquad \text{at } t = 0 \text{ for } -3 \leqslant x \leqslant 3,$$

so that the fourth-order equation has solution $u = x^4 + 24(1 - t)$. This is positive for $0 \leqslant t \leqslant 1$ and is negative in an expanding interval for $1 < t \leqslant 2$.

Note that, although Green's functions and transform methods can, in principle, be applied to parabolic equations of any order, terms like $e^{-p^{1/4}x}$ would appear when we tried, say, a Laplace transform in time on a fourth-order linear equation. This makes the inversion more unwieldy than for the heat equation.

Another class of problems leading to fourth-order parabolic equations concerns the flow of thin viscous films, such as paints or other coatings, under the action of surface tension. As usual, we denote the thickness of the film by h and its pressure by p; unlike the thin films mentioned in §6.6.1, the pressure, which in the absence of gravity is approximately equal to its value at the film surface, is now equal to the surface tension multiplied by the curvature. Hence p is approximately proportional to $-\nabla^2 h$, assuming that the film is thin and that $|\nabla h|$ is small, where ∇ is the two-dimensional gradient in the plane of the film. We can, however, still use the *lubrication theory approximation* to give that the in-plane velocity is proportional to $-h^2\nabla p$, so the mass flux in turn is proportional to $h^3\nabla^2 h$. Finally, conservation of mass now means that

$$\frac{\partial h}{\partial t} = -\nabla \cdot \left(h^3 \nabla(\nabla^2 h)\right), \tag{6.104}$$

with an appropriate scaling, and we have a fourth-order diffusion equation with cubic 'diffusivity'.

Like the Cahn–Hilliard equation, (6.104) is an equation about which a great deal more could be said. While the former is susceptible to some of the methods described in §6.6 for semilinear equations, especially the integral estimates approach, the latter clearly possesses some group invariance that can be exploited to find similarity solutions. Unfortunately, the degeneracy in (6.104) when $h = 0$ is much more acute than it was for the porous medium equation, which is a pity because painters are often interested to know criteria for the occurrence of 'pinholing', in which h tends to zero locally in space and time.[130]

In line with the remarks at the beginning of this section, we conclude with a cautionary tale about vector equations.

6.7.2 Higher-order systems

When different species of population interact, or when more than one chemical concentration is of importance in a reaction, or when simultaneous heat conduction and mass diffusion are considered, models arise in the form of coupled systems of parabolic equations. In the absence of convection, the basic form of such a system can sometimes be written as

$$\frac{\partial \mathbf{u}}{\partial t} = \nabla \cdot (\mathbf{D}\nabla \mathbf{u}) + \mathbf{f}(\mathbf{u}) \quad \text{in } \Omega,$$

where $\mathbf{u}$ and $\mathbf{f}$ are now vectors with n components, $\nabla \mathbf{u}$ is the $n \times m$ matrix $(\partial u_i/\partial x_j)$, and $\mathbf{D}$ is an $n \times n$ diagonal matrix with positive diagonal elements.[131] Each component of $\mathbf{u}$ satisfies a boundary condition on $\partial\Omega$.

One particular property of these systems is the occurrence of the so-called *Turing* or *double-diffusive instability*. Even though diffusion acts as a stabilising and smoothing mechanism for a single equation, we now have the following possibility: $\mathbf{u}$ can be a constant steady state such that $\mathbf{f}(\mathbf{u}) = \mathbf{0}$ and it can be stable as

[130] It can, however, be proved that, if h is strictly positive on the parabolic boundary Γ of Fig. 6.1, then h is strictly positive in the interior.

[131] In some thermodynamic models $\mathbf{D}$ is symmetric rather than diagonal.

the solution of the ordinary differential equation $\mathrm{d}\mathbf{u}/\mathrm{d}t = \mathbf{f}(\mathbf{u})$, so that all the eigenvalues of the Jacobian matrix $(\partial f_i/\partial u_j)$ have negative real part, yet $\mathbf{u}$ can be unstable as a solution of the parabolic system. Consider, for example, the linear system

$$\frac{\partial \mathbf{u}}{\partial t} = \begin{pmatrix} 14 & 0 \\ 0 & 1 \end{pmatrix} \frac{\partial^2 \mathbf{u}}{\partial x^2} + \begin{pmatrix} -2 & -2 \\ 2 & 1 \end{pmatrix} \mathbf{u} \quad \text{for } 0 < x < 2\pi,$$

$$\mathbf{u} = \mathbf{0} \quad \text{for } x = 0 \text{ and } x = 2\pi.$$

The system of ordinary differential equations

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \begin{pmatrix} -2 & -2 \\ 2 & 1 \end{pmatrix} \mathbf{u} = \mathbf{A}\mathbf{u}, \quad \text{say,}$$

has zero as its only steady state. Its stability is determined by the eigenvalues λ of $\mathbf{A}$; these satisfy $\lambda^2 + \lambda + 2 = 0$, so the eigenvalues have negative real parts and zero is a stable solution.

The boundary value problem has solutions of the form

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mathrm{e}^{\mu t} \sin kx$$

for $k = 1/2, 1, 3/2, \ldots$, provided that

$$(\mathbf{B} - \mu\mathbf{I})\mathbf{u} = \begin{pmatrix} -2 - 14k^2 - \mu & -2 \\ 2 & 1 - k^2 - \mu \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \mathbf{0}.$$

The eigenvalue μ now satisfies

$$\mu^2 + (1 + 15k^2)\mu + 4 - (1 - k^2)(2 + 14k^2) = 0.$$

The trace of the matrix $\mathbf{B}$ remains negative but the determinant is also negative, and so the eigenvalues are real and of opposite signs, if $(1 - k^2)(2 + 14k^2) > 4$. Taking $k = 1/2$ this inequality is indeed satisfied and the trivial solution is unstable. The effect of the coupling between the components of $\mathbf{u}$ is similar to that of the coupling terms in the system $\mathrm{d}x/\mathrm{d}t = -x + \lambda y$, $\mathrm{d}y/\mathrm{d}t = -y + \lambda x$ for $\lambda > 1$.

Despite the importance of these systems, for example in oceanic instabilities where temperature and density interact, their mathematical theory is far less well advanced than in the scalar case. Particular problems are encountered when the diffusivity matrix $\mathbf{D}$ is not diagonal and depends on $\mathbf{u}$, because of the difficulty of predicting whether or not $\mathbf{D}$ is positive definite and hence whether the system is strictly parabolic.

Exercises

6.1. Suppose a pollutant concentration c satisfies

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = D\nabla^2 c, \quad v = \text{constant},$$

with

$$\frac{\partial c}{\partial y} = 0 \quad \text{on } y = \pm d$$

and

$$c = c_0 \quad \text{at } x = 0, \qquad c = c_1 \quad \text{at } x = L.$$

Write $x = LX$, $y = dY$ and $t = (L/v)\tau$ to obtain

$$\frac{\partial c}{\partial \tau} + \frac{\partial c}{\partial X} = k\left(\frac{\partial^2 c}{\partial Y^2} + \delta^2 \frac{\partial^2 c}{\partial X^2}\right), \tag{6.105}$$

where $k = DL/vd^2$ and $\delta = d/L$. Show there is a steady solution $c = c(X)$ which tends to c_0 as $\delta \to 0$ except near $X = 1$.
If $\delta \ll 1$ and the last term in (6.105) is neglected, which boundary conditions would you expect to be able to satisfy?

6.2. Find the time-periodic solution of

$$\rho c\frac{\partial T}{\partial t} = k\frac{\partial^2 T}{\partial x^2} \quad \text{for } x > 0 \qquad \text{with} \qquad -k\left.\frac{\partial T}{\partial x}\right|_{x=0} = Q\cos\omega t$$

and T bounded as $x \to \infty$. Take $\omega = 2\pi\,(\text{year})^{-1}$ to give a reason why the hottest (coldest) day of the year might be expected to occur about six weeks after the longest (shortest) day.

6.3. Suppose w is the difference between two solutions of

$$\frac{\partial u}{\partial t} = \nabla^2 u + au + f(x,t) \quad \text{in } \Omega,$$

with

$$\frac{\partial u}{\partial n} + \alpha u = g(x,t) \quad \text{on } \partial\Omega$$

and

$$u = h(x) \quad \text{at } t = 0,$$

where a and α are constants and $\alpha > 0$. Show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega w^2\,\mathrm{dx} + \frac{1}{2}\int_\Omega \left(|\nabla w|^2 - aw^2\right)\mathrm{dx} + \frac{\alpha}{2}\int_{\partial\Omega} w^2\,\mathrm{d}S = 0.$$

Deduce that

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \text{constant}\right)\int_\Omega w^2\,\mathrm{dx} \leqslant 0,$$

and thus that $w \equiv 0$.

*6.4. Suppose

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \quad \text{in } -L < x < \infty,$$

with

$$u(x,0) = 0, \qquad u(-L,t) = u_0 = \text{constant}, \qquad u \to 0 \quad \text{as } x \to +\infty.$$

Show that there is a solution

$$u(x,t) = \frac{2u_0}{\sqrt{\pi}} \int_{(x+L)/\sqrt{t}}^{\infty} e^{-\eta^2/4}\, d\eta.$$

Since $\int_X^\infty e^{-\eta^2/4}\, d\eta < e^{-X^2/4}$ for large *positive* X, if $u_0 = e^{KL^2}$ and $L \gg 1$, then

$$u < \frac{2}{\sqrt{\pi}} e^{KL^2 - (x+L)^2/4t}$$

unless x is large and negative. Deduce that, if the boundary condition is replaced by $u < e^{Kx^2}$ for all x sufficiently large and negative, and some $K > 0$, then $u \equiv 0$ for t sufficiently small.

6.5. Show that, if $f(t)$ is infinitely differentiable, then

$$u(x,t) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \frac{d^n f(t)}{dt^n}$$

satisfies the heat equation. Deduce that there are non-zero solutions of the heat equation that satisfy $u(x,0) = 0$ for all x.

6.6. A particle moves along the real line, starting at the origin and taking a step of $\pm h$ with equal probability $1/2$ in each time step k. If $p(x,t)$ is the probability density function of its position at time t, use the argument of §6.1.2 to show that, in the limit $h, k \to 0$ with $h^2 = k$,

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}.$$

Explain why the appropriate initial condition is $p(x,0) = \delta(x)$, and show that $p(x,t) = (1/\sqrt{2\pi t})e^{-x^2/2t}$, the probability density of a normal distribution with mean zero and variance t.

Remark. This is the density function of Brownian motion. The discrete random walk (before taking the limit $h, k \to 0$) has a translated binomial distribution, and the normality of the limiting distribution is an example of the central limit theorem; it can also be established by taking the limit of the binomial probabilities.

6.7. It is observed numerically that

$$\left(\sum_{-\infty}^{\infty} e^{-n^2}\right)^2 \approx 3.142.$$

Use a periodic solution of the one-dimensional heat equation to prove that $\sum_{-\infty}^{\infty} e^{-n^2} > \sqrt{\pi}$.

6.8. Let $G(x,t;\xi)$ satisfy

$$\frac{\partial G}{\partial t} = \frac{\partial^2 G}{\partial x^2} \quad \text{for } 0 < x < 1,\ t > 0,$$

with

$$\frac{\partial G}{\partial x} = 0 \quad \text{at } x = 0,1 \qquad \text{and} \qquad G = \delta(x-\xi) \quad \text{at } t = 0.$$

Show that the Laplace transform of G in time satisfies

$$\frac{\mathrm{d}^2\tilde{G}}{\mathrm{d}x^2} - p\tilde{G} = \delta(x-\xi), \qquad \frac{\mathrm{d}\tilde{G}}{\mathrm{d}x} = 0 \quad \text{at } x = 0,1,$$

and hence that

$$\tilde{G}(x,\xi;p) = \begin{cases} \cosh\left(x\sqrt{p}\right)\cosh\left((1-\xi)\sqrt{p}\right) / \left(\sqrt{p}\sinh\sqrt{p}\right), & 0 \leqslant x \leqslant \xi, \\ \cosh\left(\xi\sqrt{p}\right)\cosh\left((1-x)\sqrt{p}\right) / \left(\sqrt{p}\sinh\sqrt{p}\right), & \xi \leqslant x \leqslant 1. \end{cases}$$

Deduce that

$$\begin{aligned} G(x,t;\xi) &= \frac{1}{2\pi\mathrm{i}} \int_{\Re\, p=\text{constant}>0} \tilde{G}\mathrm{e}^{pt}\,\mathrm{d}p \\ &= \begin{cases} \sum_{-\infty}^{\infty} (-1)^n \mathrm{e}^{-n^2\pi^2 t} \cos n\pi x \cos n\pi(1-\xi), & x \leqslant \xi, \\ \sum_{-\infty}^{\infty} (-1)^n \mathrm{e}^{-n^2\pi^2 t} \cos n\pi\xi \cos n\pi(1-x), & \xi \leqslant x, \end{cases} \end{aligned}$$

and hence that

$$G(x,t;\xi) = 1 + 2\sum_{n=1}^{\infty} \mathrm{e}^{-n^2\pi^2 t} \cos n\pi x \cos n\pi\xi.$$

Remarks.

(i) The Green's function for this problem is $G(x,\tau-t;\xi)$. The function G here can be identified with G' on p. 251.

(ii) G can be written as

$$\frac{1}{2}\left(\theta_3\left(\frac{x-\xi}{2\pi}, \mathrm{e}^{-\pi^2 t}\right) - \theta_3\left(\frac{x+\xi}{2\pi}, \mathrm{e}^{-\pi^2 t}\right)\right),$$

where θ_3 is a *theta function.*

(iii) This result can be related to (6.36) by expanding $\tilde{G}$ as a power series in $\mathrm{e}^{-2\sqrt{p}}$ and inverting term by term.

6.9. If

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } t > 0, \qquad \frac{\partial u}{\partial x} = 0 \quad \text{at } x = 0,1, \qquad u(x,0) = u_0(x),$$

use the result that

$$u(\xi,\tau) = \int_0^1 u_0(x) G(x,\tau;\xi)\,\mathrm{d}x,$$

where G is given in Exercise 6.8, to show that $u(\xi,\tau) \to \int_0^1 u_0(x)\,\mathrm{d}x$ as $\tau \to \infty$.

6.10. Suppose the Neumann condition in Exercise 6.9 is replaced by the Dirichlet condition $G = 0$ at $x = 0, 1$. Show that G becomes

$$2\sum_{n=1}^{\infty} e^{-n^2\pi^2 t} \sin n\pi x \sin n\pi\xi$$
$$= \frac{1}{2\sqrt{\pi t}} \sum_{m=-\infty}^{\infty} \left(e^{-(x-\xi-2m)^2/4t} - e^{-(x+\xi-2m)^2/4t}\right).$$

6.11. Suppose that $f(x)$ is periodic with period 2π and has the Fourier series

$$f(x) = \sum_{n=N}^{\infty} (a_n \cos nx + b_n \sin nx),$$

with the first $N - 1$ harmonics absent. Use the large-t behaviour of the periodic solution of $\partial u/\partial t = \partial^2 u/\partial x^2$ for $t > 0$, with $u(x, 0) = f(x)$, to show that, for large t, $u(x, t)$ has at least $2N$ zeros in each period. Deduce from the maximum principle that, as t increases, zeros of u can disappear but not appear, and hence show that $f(x)$ has at least $2N$ zeros in each period.

6.12. Show that the Green's function for $\partial u/\partial t = \partial^2 u/\partial x^2$ in $x > 0$, with u prescribed at $x = 0$, is, for $t < \tau$,

$$G(x, \tau - t; \xi) = \frac{1}{2\sqrt{\pi(\tau - t)}} \left(e^{-(x-\xi)^2/4(\tau-t)} - e^{-(x+\xi)^2/4(\tau-t)}\right).$$

Using the result that

$$\int_0^\tau \int_0^\infty \left(\left(\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t}\right) G - \left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial G}{\partial t}\right) u\right) dx\, dt = u(\xi, \tau),$$

show that, if $u = g(t)$ on $x = 0$ and $u = h(x)$ at $t = 0$, then

$$u(\xi, \tau) = \frac{1}{2\sqrt{\pi\tau}} \int_0^\infty h(x) \left(e^{-(x-\xi)^2/4\tau} - e^{-(x+\xi)^2/4\tau}\right) dx$$
$$+ \frac{\xi}{2\sqrt{\pi}} \int_0^\tau \frac{g(t) e^{-\xi^2/4(\tau-t)}}{(\tau - t)^{3/2}}\, dt.$$

6.13. Suppose $w(x, t)$ has Laplace transform $\widetilde{w} = e^{-\sqrt{p}x}$, $\Re\sqrt{p} > 0$. By deforming the inversion contour to lie along the negative real axis, show that, for $\gamma > 0$,

$$w(x, t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{pt - \sqrt{p}x}\, dp = \frac{1}{\pi} \int_0^\infty e^{-st} \sin\left(x\sqrt{s}\right) ds.$$

Deduce that

$$w(x, t) = \frac{-2}{\pi\sqrt{t}} \frac{\partial}{\partial x} \int_0^\infty e^{-y^2} \cos\left(\frac{xy}{\sqrt{t}}\right) dy = \frac{x}{2(\pi t)^{3/2}} e^{-x^2/4t}.$$

6.14. By taking Laplace transforms, show that the solution of

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{in } x > 0,\ t > 0,$$

with

$$u = 0 \qquad \text{at } t = 0$$

and

$$u = g(t) \quad \text{on } x = 0,$$

is given by the *Duhamel formula*

$$u(x,t) = \int_0^t g(t-\tau)\frac{\partial v}{\partial t}(x,\tau)\,\mathrm{d}\tau,$$

where

$$\frac{\partial^2 v}{\partial x^2} = \frac{\partial v}{\partial t} \quad \text{in } x > 0,\ t > 0, \qquad v(x,0) = 0, \qquad v(0,t) = 1.$$

How is v related to the Green's function for this problem? What is the physical interpretation of the formula for u?

*6.15. Show that, if

$$\nabla^2 G = \frac{\partial G}{\partial t} \quad \text{in } \Omega,\ t > 0,$$

with

$$G = 0 \quad \text{on } \partial\Omega \qquad \text{and} \qquad G = \delta(\mathbf{x} - \boldsymbol{\xi}) \quad \text{at } t = 0,$$

then the eigenvalues $\lambda_0, \lambda_1, \ldots$ of the Helmholtz problem

$$(\nabla^2 + \lambda)\phi = 0 \quad \text{in } \Omega, \qquad \phi = 0 \quad \text{on } \partial\Omega$$

are such that

$$\sum_{n=0}^{\infty} \mathrm{e}^{-\lambda_n t} = \int_\Omega G(\mathbf{x}, t; \mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Assuming you can approximate G for small t by

$$\frac{1}{(4\pi t)^{m/2}}\mathrm{e}^{-|\mathbf{x}-\boldsymbol{\xi}|^2/4t},$$

where m is the dimension of Ω, show that, as $t \to 0$, $\sum_{n=0}^{\infty} \mathrm{e}^{-\lambda_n t}$ is approximated by $\mathrm{vol}(\Omega)/(4\pi t)^{m/2}$.

Remark. To make this rigorous, you need to show that the contribution from the region near the boundary, where the approximation for G is invalid, is negligibly small as $t \to 0$.

6.16. Show that the Green's function for

$$\frac{\partial u}{\partial t} + (\mathbf{v} \cdot \nabla)u = \nabla^2 u \quad \text{in } \Omega,$$

where $\mathbf{v}$ is prescribed, and where u is given on $\partial\Omega$ and at $t = 0$, satisfies

$$\frac{\partial G}{\partial t} + \nabla \cdot (G\mathbf{v}) + \nabla^2 G = \delta(\mathbf{x} - \boldsymbol{\xi})\delta(t - \tau),$$

with $G = 0$ on $\partial\Omega$ and at $t = T > \tau$. Hence show that, if Ω is the real line $-\infty < x < \infty$ and $\mathbf{v}$ is a constant vector in the x direction, then, for $t < \tau$,

$$G(x, \tau - t; \xi) = \frac{1}{2\sqrt{\pi(\tau - t)}} e^{-(x-\xi+|\mathbf{v}|(\tau-t))^2/4(\tau-t)}.$$

6.17. Verify that the similarity solution of

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \text{in } x > 0,\, y > 0,\, t > 0,$$

with $u = 1$ at $t = 0$, and $u = 0$ on $x = 0$ and $y = 0$, is

$$u = \operatorname{erf}\left(\frac{x}{2\sqrt{t}}\right) \operatorname{erf}\left(\frac{y}{2\sqrt{t}}\right),$$

where $\operatorname{erf} \eta = (2/\sqrt{\pi}) \int_0^\eta e^{-s^2}\, ds$.

6.18. Suppose that

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial y^2} \quad \text{in } t > 0,\, x > 0,\, y > 0,$$

and $u(0, y, t) = u(x, y, 0) = 0$ and $u(x, 0, t) = 1$. Would you expect u to be an analytic function of x and t? Show that

$$u = \begin{cases} \operatorname{erfc}\left(y/2\sqrt{t}\right), & 0 < t < x, \\ \operatorname{erfc}\left(y/2\sqrt{x}\right), & x < t, \end{cases}$$

where $\operatorname{erfc} \eta = 1 - \operatorname{erf} \eta$, satisfies the equations and boundary conditions except at $x = t$. What discontinuity occurs at $x = t$?

*6.19. Suppose that

$$D\frac{d^2\psi}{dy^2} - (p - \lambda v(y))\,\psi = 0 \quad \text{for } v > 0,$$

with

$$\frac{d\psi}{dy} = 0 \quad \text{at } y = 0, 1.$$

For small p, write $\psi = \psi_{(0)} + p\psi_{(1)} + p^2\psi_{(2)} + \cdots$ and $\lambda = p\lambda_{(1)} + p^2\lambda_{(2)} + \cdots$ and equate powers of p to show that

$$\psi_{(0)} = c_0 = \text{constant}, \qquad \lambda_{(1)} = \frac{1}{v_0}, \quad \text{where} \quad v_0 = \int_0^1 v(y)\,\mathrm{d}y,$$

$$\psi_{(1)} = \frac{c_0}{Dv_0}\int_0^y (y - y')\,(v_0 - v(y'))\,\mathrm{d}y' + \text{constant}$$

and

$$D\frac{\mathrm{d}^2\psi_{(2)}}{\mathrm{d}y^2} = \left(1 - \frac{v}{v_0}\right)\psi_{(1)} - \lambda_{(2)}c_0v_0.$$

Deduce that

$$\lambda_{(2)} = \frac{1}{c_0v_0}\int_0^1 \left(1 - \frac{v(y)}{v_0}\right)\psi_{(1)}\,\mathrm{d}y = -\frac{D_0}{v_0^3},$$

say, where

$$D_0 = \frac{v_0^2}{D}\int_0^1 \left[\int_0^y \left(1 - \frac{v(y')}{v_0}\right)\mathrm{d}y'\right]^2 \mathrm{d}y \geqslant 0.$$

6.20. (i) Suppose $x' = g(x,\lambda)$ and $x'' = g(x',\mu)$, where g is a group and $g(x,0) = x$. By differentiating with respect to μ and setting $\mu = 0$, show that $\partial g/\partial\lambda$ is the product of a function of g and a function of λ.

(ii) Suppose

$$\frac{\partial g}{\partial\lambda}(x,\lambda) = F(g(x,\lambda))$$

and $g(x,0) = x$. Show that there is a function G such that $g = G(\lambda + G^{-1}(x))$ and hence verify the closure condition for $x \mapsto g(x,\lambda)$ to form a group.

*6.21. Let (6.56) be generalised to

$$x' = f(x,t;\lambda), \quad t' = g(x,t;\lambda), \quad u' = h(x,t,u;\lambda)$$

and

$$\mathcal{U} = U\frac{\partial}{\partial x} + V\frac{\partial}{\partial t} + W\frac{\partial}{\partial u};$$

when u is replaced by u' in (6.59), and the $O(\lambda)$ terms are collected on the left-hand side, show that there is a new term

$$\lambda\left(-\frac{\partial W}{\partial u}\frac{\partial u'}{\partial t'} - \frac{\partial W}{\partial t} + \frac{\partial W}{\partial u}\frac{\partial^2 u'}{\partial x'^2} + 2\frac{\partial^2 W}{\partial x\partial u}\frac{\partial u'}{\partial x'} + \frac{\partial^2 W}{\partial u^2}\left(\frac{\partial u'}{\partial x'}\right)^2 + \frac{\partial^2 W}{\partial x^2}\right).$$

Deduce that the heat equation is invariant as long as

$$\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2} + 2\frac{\partial U}{\partial x}, \quad \frac{\partial V}{\partial x} = 0,$$

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2} - 2\frac{\partial^2 W}{\partial x\partial u}, \quad \frac{\partial W}{\partial t} = \frac{\partial^2 W}{\partial x^2}, \quad \frac{\partial^2 W}{\partial u^2} = 0.$$

Verify that (6.68) satisfies these equations and deduce that one possibility is $f = e^{\lambda/2}x$, $g = e^{\lambda}t$ and $h = e^{\gamma\lambda}u$, where γ is an arbitrary constant, and that ξ is invariant under the transformation group if

$$\left(\frac{x}{2}\frac{\partial}{\partial x} + t\frac{\partial}{\partial t} + \gamma u\frac{\partial}{\partial u}\right)\xi = 0.$$

Finally, show that $x/\sqrt{t}$ and u/t^{γ} are both invariant, so that a solution of the heat equation is $u = t^{\gamma}F(x/\sqrt{t})$, where

$$\frac{d^2F}{d\eta^2} + \frac{\eta}{2}\frac{dF}{d\eta} - \gamma F = 0.$$

6.22. Show that the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\nabla^2 u$$

is *not* invariant under the Galilean group of transformations

$$x' = x - \lambda t, \quad y' = y, \quad z' = z, \quad t' = t$$

(for which the group operation is $\nu = \lambda + \mu$ in the notation of p. 266), where the constant λ, with $|\lambda| < c$, represents the speed of the (x', y', z') coordinate system along the x axis. Show that it is, however, invariant under the *Lorentz transformation*

$$x' = \gamma(x - \lambda t), \quad y' = y, \quad z' = z, \quad t' = \gamma\left(t - \frac{\lambda x}{c^2}\right),$$

where $\gamma = 1/\sqrt{1 - \lambda^2/c^2}$, and that these transformations form a group with group operation $\nu = (\lambda + \mu)/(1 + \lambda\mu/c^2)$.

Remark. Recalling that the components of the electric and magnetic fields in a vacuum satisfy the wave equation above, this invariance is consistent with the fact that the speed of light c is the same in all inertial frames, the basic postulate of the theory of special relativity. It is also possible to show that a suitable formulation of Maxwell's equations is invariant under Lorentz transformations. The group operation of the Lorentz transformations corresponds to the relativistic formula for the addition of velocities.

6.23. The equation

$$x\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } x > 0,\ t > 0$$

models advection in a viscous flow, where x is the distance from a solid boundary and t is the distance downstream. Show that the equation is invariant when we set

$$x' = e^{\lambda/3}x, \quad t' = e^{\lambda}t, \quad u' = e^{\gamma\lambda}u.$$

Deduce that $t^{\gamma}F(x/t^{1/3})$ is a solution as long as

$$\frac{\mathrm{d}^2F}{\mathrm{d}\eta^2}+\frac{\eta^2}{3}\frac{\mathrm{d}F}{\mathrm{d}\eta}-\gamma\eta F=0,$$

and that one possibility is $u=t^{-2/3}\exp(-x^3/9t)$.

6.24. Show that, if

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$$

is invariant under $x'=f(x,y;\lambda)$, $y'=g(x,y;\lambda)$, then the components U and V of the infinitesimal generator satisfy

$$\frac{\partial U}{\partial x}=\frac{\partial V}{\partial y},\quad \frac{\partial V}{\partial x}=-\frac{\partial U}{\partial y}.$$

6.25. Suppose that $\nabla^2 u=0$, with

$$u=\begin{cases}0, & \theta=0,\ r<1,\\ r^2, & \theta=\pi/2,\ r<1,\\ f(\theta), & r=1,\end{cases}$$

in two-dimensional polar coordinates. Show that

$$u=-\frac{2}{\pi}r^2(\theta\cos 2\theta+\log r\sin 2\theta)+\sum_{n=1}^{\infty}b_n r^{2n}\sin 2n\theta,$$

where

$$f(\theta)+\frac{2}{\pi}\theta\cos 2\theta=\sum_{n=1}^{\infty}b_n\sin 2n\theta\quad\text{for } 0<\theta<\frac{\pi}{2}.$$

(This question generalises the result of Exercise 5.12.)

*6.26. Suppose that

$$\frac{\partial u}{\partial t}=\frac{1}{\epsilon}\nabla^2 u+f(u)\quad\text{in }\Omega,$$

with

$$\frac{\partial u}{\partial n}+\epsilon u=0\quad\text{on }\partial\Omega.$$

Show that, if you write $u=u_0+\epsilon u_1+\cdots$ for small ϵ and formally equate powers of ϵ, you obtain

$$\nabla^2 u_0=0,\qquad \frac{\partial u_0}{\partial n}=0\quad\text{on }\partial\Omega,$$

so that $u_0=u_0(t)$. Then show that

$$\nabla^2 u_1=\frac{\mathrm{d}u_0}{\mathrm{d}t}-f(u_0)\qquad\text{with}\qquad \frac{\partial u_1}{\partial n}=-u_0\quad\text{on }\partial\Omega,$$

and deduce that

$$\frac{\mathrm{d}u_0}{\mathrm{d}t}-f(u_0)=u_0\frac{\mathrm{area}(\partial\Omega)}{\mathrm{vol}(\Omega)}.$$

*6.27. The system

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0,$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) + \frac{\partial p}{\partial x} = \epsilon\frac{\partial^2 u}{\partial x^2},$$

$$\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + \gamma p\frac{\partial u}{\partial x} = \epsilon(\gamma - 1)\left(\frac{\partial u}{\partial x}\right)^2,$$

with ϵ and γ constant, models viscous effects in one-dimensional gas dynamics (the right-hand sides of the second and third equations describe the resistance to extension (as in (2.11)) and the rate of working of this resistance, respectively). Suppose that u, $p-p_0$ and $\rho-\rho_0$ are all small, where p_0 and ρ_0 are constant. Then the flow is localised about waves travelling with speed $dx/dt = \pm a_0$, where $a_0^2 = \gamma p_0/\rho_0$. To study a wave near $x = a_0 t$ in detail, write $x = a_0 t + \delta y$; when terms of order ϵ, ϵ^2/δ and δ are neglected, show that the perturbations satisfy

$$-a_0\frac{\partial \bar\rho}{\partial y} + \rho_0\frac{\partial \bar u}{\partial y} = 0, \quad -\rho_0 a_0\frac{\partial \bar u}{\partial y} + \frac{\partial \bar p}{\partial y} = 0, \quad -a_0\frac{\partial \bar p}{\partial y} + \gamma p_0\frac{\partial \bar u}{\partial y} = 0.$$

The linear dependence of these equations means that we must retain terms quadratic in the perturbation along with terms involving ϵ and δ. We only get a sensible answer when $\epsilon = \lambda\delta^2$ for some $O(1)$ constant λ. Show that, with these assumptions, the largest terms to be retained on the right-hand side of the three equations are

$$-\delta\frac{\partial \bar\rho}{\partial t} - \bar u\frac{\partial \bar\rho}{\partial y} - \bar\rho\frac{\partial \bar u}{\partial y}, \quad -\delta\rho_0\frac{\partial \bar u}{\partial t} - \rho_0\bar u\frac{\partial \bar u}{\partial y} + \delta\frac{\partial^2 \bar u}{\partial y^2} + a_0\bar\rho\frac{\partial \bar u}{\partial y},$$
$$-\delta\frac{\partial \bar p}{\partial t} - \bar u\frac{\partial \bar p}{\partial y} - \gamma\bar p\frac{\partial \bar u}{\partial y},$$

respectively. Finally, use cross-differentiation to show that

$$-2\rho_0\frac{\partial \bar u}{\partial t} + \lambda\frac{\partial^2 \bar u}{\partial y^2} - \rho_0\frac{\gamma+1}{\delta}\bar u\frac{\partial \bar u}{\partial y} = 0,$$

which is (2.50) with a change of notation.

6.28. Suppose that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u(1-u)$$

and $u = F(x - Vt)$, with $V > 0$. Show that

$$\frac{d^2F}{d\xi^2} + V\frac{dF}{d\xi} + F(1-F) = 0,$$

and that the (F, F') phase-plane has a saddle point at $(1,0)$, and a stable node or spiral at $(0,0)$, depending whether $V > 2$ or $V < 2$, respectively.

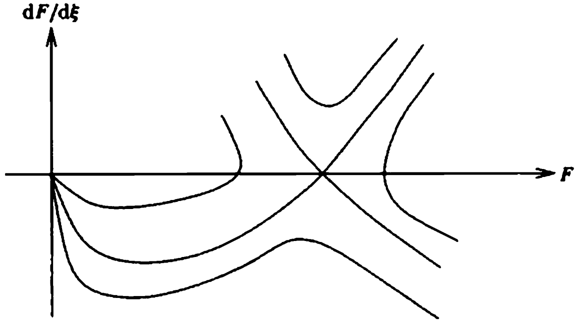


Fig. 6.5 Phase-plane for travelling waves of the Fisher equation.

Show that the phase-plane is as shown schematically in Fig. 6.5 and deduce that there is a monotone F in which $F \to 0$ as $x - Vt \to +\infty$ and $F \to 1$ as $x - Vt \to -\infty$, as long as $V \geqslant 2$.

6.29. Show that travelling wave solutions of

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u(1-u)(u-a) \quad \text{for } 0 < a < \frac{1}{2}$$

satisfy

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\xi^2} + V\frac{\mathrm{d}F}{\mathrm{d}\xi} + F(1-F)(F-a) = 0.$$

Using the fact that the phase-plane of $(F, \mathrm{d}F/\mathrm{d}\xi)$ for $a < F < 1$ is similar to that of Fig. 6.5 for $0 < F < 1$, show that travelling waves with $F(-\infty) = a$ and $F(\infty) = 1$ exist if

$$V \geqslant V_{\min} \quad \text{for some } V_{\min} \geqslant 2\sqrt{a(1-a)}.$$

By considering a sequence of phase-planes as V decreases, show that there is a unique value of $V < V_{\min}$ such that there is a travelling wave with $F(-\infty) = 0$ and $F(\infty) = 1$, and that this value is zero when $a = \frac{1}{2}$.

6.30. Suppose that

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{1}{u^2}\frac{\partial u}{\partial x}\right).$$

Define ψ to satisfy

$$\frac{\partial \psi}{\partial x} = u, \quad \frac{\partial \psi}{\partial t} = \frac{1}{u^2}\frac{\partial u}{\partial x},$$

and make the partial hodograph transformation from (x, t) to (ψ, t) to show that $v = 1/u$ satisfies the heat equation

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial \psi^2}.$$

6.31. Blasius' equation,

$$\frac{\mathrm{d}^2F}{\mathrm{d}\eta^3} + \frac{1}{2}F\frac{\mathrm{d}^2F}{\mathrm{d}\eta^2} = 0,$$

is autonomous (i.e. invariant under $\eta' = \eta + \lambda$), so we can lower its order by writing $F = G$ and $\mathrm{d}F/\mathrm{d}\eta = H$ to give

$$H\frac{\mathrm{d}^2H}{\mathrm{d}G^2} + \left(\frac{\mathrm{d}H}{\mathrm{d}G}\right)^2 + \frac{1}{2}G\frac{\mathrm{d}H}{\mathrm{d}G} = 0.$$

This equation is itself invariant under $G' = (1+\lambda)G$, $H' = (1+\lambda)^2H$, for which the infinitesimal generator is

$$\mathcal{U} = G\frac{\partial}{\partial G} + 2H\frac{\partial}{\partial H}.$$

Show that $\mathcal{U}X = 1$ and $\mathcal{U}Y = 0$ are satisfied by $Y = H/G^2$ and $X = \log G$, and derive an autonomous second-order equation for $Y(X)$.

6.32. Suppose that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \mathrm{e}^u \quad \text{for } x > 0,$$

with

$$u = 0 \quad \text{on } x = 0, \qquad u = 0 \quad \text{at } t = 0.$$

Show that $-\log(1-t)$ is an upper solution and $-\log(1 - t + h(x,t))$ is a lower solution as long as $\partial h/\partial t = \partial^2 h/\partial x^2$, with $h = t$ on $x = 0$ and $h = 0$ at $t = 0$. Deduce that $u(x,1) \to \infty$ as $x \to \infty$.

*6.33. Suppose that a smooth plane curve $\mathbf{r} = \mathbf{r}(s,t)$, where s is the arc length from any prescribed point on the curve and t is time, evolves by a curvature flow in which its velocity v_n along its normal $\mathbf{n}$ is equal to its curvature $\kappa(s,t)$. Defining the unit tangent by $\mathbf{t} = \partial\mathbf{r}/\partial s$, show that

$$\frac{\partial \mathbf{r}}{\partial t} = \kappa\mathbf{n} + u(s,t)\mathbf{t}$$

for some function $u(s,t)$. Use the Serret–Frenet formulæ

$$\frac{\partial \mathbf{t}}{\partial s} = \kappa\mathbf{n}, \qquad \frac{\partial \mathbf{n}}{\partial s} = -\kappa\mathbf{t}$$

to show that

$$\frac{\partial \kappa}{\partial t} = \frac{\partial^2 \kappa}{\partial s^2} + \kappa^3 + u\frac{\partial \kappa}{\partial s}, \qquad \frac{\partial u}{\partial s} - \kappa^2 = 0.$$

Show that, if the curve is simple and closed and has length $L(t)$, so that $\int_0^L \kappa\,\mathrm{d}s = 2\pi$, then

$$\kappa(L(t),t)\frac{\mathrm{d}L}{\mathrm{d}t} = -\left[\kappa L\right]_0^L = -\kappa(L(t),t)\left[u\right]_0^L.$$

Deduce that

$$\frac{dL}{dt} = -\int_0^{L(t)} \kappa^2(s,t)\, ds,$$

so that the curve grows shorter. If the curve encloses a region Ω, show that the rate of change of area of Ω is $\int_{\partial\Omega} v_n\, ds$, and hence that the area of Ω decreases at the rate 2π.

If two simple closed curves evolve by curvature flow, and one is initially inside the other, show that it remains inside (consider what would happen if they were to touch, the inside curve having larger curvature at the point of contact). Show that the curve $e^{-(x-Vt)} = \cos \pi y$, $V > 0$, is a travelling wave solution of (6.88). This curve, known as the Grim Reaper because any other curves in its path vanish in finite time by the comparison result just shown, also occurs as the free boundary for a famous Hele-Shaw flow in a parallel-sided channel, called a *Saffman–Taylor finger* (see Exercise 7.19).

6.34. Suppose that

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u) \quad \text{in } \Omega \qquad \text{with} \qquad u = 0 \quad \text{on } \partial\Omega,$$

that $f > 0$, and that $f(u)/u$ is a decreasing function for $u > 0$. Suppose also that there is a positive steady state u_0 and that $u - u_0 = w$ is small. Show that, approximately,

$$\frac{\partial w}{\partial t} = \nabla^2 w + f'(u_0)w.$$

Now suppose that λ_0 is the principal eigenvalue of

$$\nabla^2\phi + (f'(u_0) + \lambda)\phi = 0, \qquad \phi = 0 \quad \text{on } \partial\Omega,$$

and that μ_0 is the principal eigenvalue of

$$\nabla^2\phi + \left(\frac{f(u_0)}{u_0} + \mu\right)\phi = 0, \qquad \phi = 0 \quad \text{on } \partial\Omega.$$

Using the appropriate generalisation of the Rayleigh quotient to characterise λ_0 and μ_0, show that (i) $\mu_0 \leqslant \lambda_0$ and (ii) $\mu_0 = 0$. Deduce that u_0 is linearly stable.

6.35. Suppose that

$$\frac{\partial u}{\partial t} = \nabla^2 u + \lambda f(u)$$

in Ω, with $u = 0$ on $\partial\Omega$ and $u = g$ at $t = 0$, where f is convex and positive. Suppose also that ϕ_0 is the principal eigenfunction of $\nabla^2\phi + \mu\phi = 0$ in Ω, $\phi = 0$ on $\partial\Omega$, with corresponding eigenvalue μ_0 and with $\int \phi_0\, d\mathbf{x} = 1$. Use Jensen's inequality to show that $a(t) = \int_\Omega u\phi_0\, d\mathbf{x}$ satisfies

$$\frac{da}{dt} \geqslant \lambda f(a) - \mu_0 a \quad \text{with} \quad a(0) = \int_\Omega \phi_0 g\, d\mathbf{x}.$$

Deduce that u blows up in finite time if

$$\int_{a(0)}^{\infty} \frac{da}{\lambda f(a) - \mu_0 a}$$

is bounded.

*6.36. Suppose that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } 0 < x < 1,\ t > 0 \qquad \text{with} \qquad u(x,0) = 0$$

and *either* one boundary condition is prescribed at each end of the interval $0 < x < 1$,

$$u(0,t) = u_0(t), \quad u(1,t) = u_1(t),$$

or two boundary conditions are prescribed at $x = 0$ and none at $x = 1$,

$$u(0,t) = u_0(t), \quad \frac{\partial u}{\partial x}(0,t) = v_0(t).$$

Solve formally by taking a Laplace transform in t and, by considering the convergence of the inverse transform, show that the problem is well posed with the first conditions but not with the second.
Repeat for the equation $\partial u/\partial t = \partial^3 u/\partial x^3$ with all permutations of boundary conditions at $x = 0$ and $x = 1$. How does your result compare with the case $\partial u/\partial t = -\partial^3 u/\partial x^3$?

6.37. Consider the system

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + au + bv, \quad \frac{\partial v}{\partial t} = D\frac{\partial^2 v}{\partial x^2} + cu + dv,$$

where a, b, c, d and D are constants, and suppose that

$$a + d < 0 < ad - bc,$$

so that the solution $u = v = 0$ is asymptotically stable when diffusion is neglected ($\partial/\partial x = 0$). Show that there are solutions in which u and v are proportional to $e^{\lambda t} \sin kx$, with $\Re \lambda > 0$, if

$$Dk^4 - (d + aD)k^2 + ad - bc < 0;$$

show that this is possible only if $ad < 0$, $bc < 0$, and, if a (or d) is negative, D is less than (or greater than) the smaller root of

$$(az - d)^2 + 4bcz = 0.$$

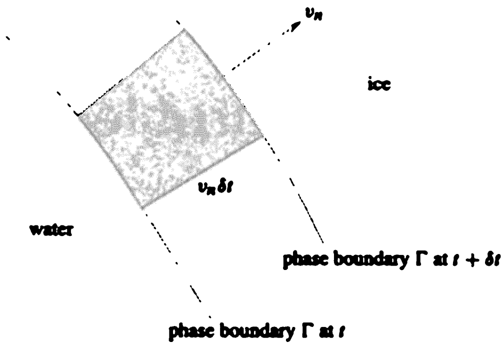


Fig. 7.1 The Stefan condition.

discussed in the previous three chapters, so, in each case, we will focus attention on what happens at the free boundary.

7.1.1 Stefan and related problems

We begin with this, the most famous free boundary problem for parabolic equations, because of its importance in subjects ranging from metal making to option pricing. In its simplest guise, the Stefan problem arises as a model for a continuous medium that can transfer heat solely by conduction, so that, as in (6.1), the temperature u satisfies the heat equation

$$\rho c\frac{\partial u}{\partial t} = k\nabla^2 u. \tag{7.1}$$

The crucial new ingredient is that we allow the material to change phase, for example to melt, freeze, vaporise or condense, at a temperature $u = u_m$ which we assume to be a given constant. Thus we need to solve (7.1), with suitable initial and fixed boundary conditions, on either side of a free boundary, the phase boundary. A simple conservation of heat at this boundary is illustrated in Fig. 7.1 for melting. For a small area A of ice, say, this gives

$$\rho L A v_n\,\delta t = [-k\nabla u\cdot\mathbf{n}]_{\text{solid}}^{\text{liquid}}\,A\,\delta t, \tag{7.2}$$

where v_n is the velocity of the free boundary normal to itself and L is the latent heat per unit mass that needs to be supplied to the ice at $u = u_m - 0$ to convert it to water at $u = u_m + 0$. Hence, in addition to the Dirichlet condition

$$u \to u_m \tag{7.3}$$

as we approach the free boundary Γ from either side, we also have the *Stefan condition* on the normal derivatives,[133]

[133]This 'box' argument is analogous to that noted in §1.7 as an alternative way of deriving the Rankine–Hugoniot conditions.

7

Free boundary problems

7.1 Introduction and models

This chapter is the most unconventional in the book. Whereas hyperbolic, elliptic and parabolic problems have been studied over many decades, and many texts are devoted to each, the subject of free boundary problems has attracted few specialised publications despite its importance in modern applied mathematics. In fact, we have already encountered several examples of this type of problem, although we have not classified them as such. For example, the thin film flows described in Chapter 1, the shock waves in Chapter 2, the contact problem in Chapter 5, and the porous medium equation of Chapter 6 could all be posed as problems involving the solution of partial differential equations in domains that are unknown *a priori*; not only the solution of the equations but also the domain of definition of the equations must be determined. We were able, by judicious approximation, to reduce the film flow problem to a differential equation for the shape of the free boundary, which in this case is the surface of the film, but all the other examples have the distinctive attribute that the free boundary geometry must be calculated and the field equations must be solved simultaneously. We will adopt this as our working definition of free boundary problems[132] and we note immediately that they are inevitably nonlinear, because the solutions of partial differential equations almost never depend linearly on the geometry of the boundaries within which they are to be solved.

We should naturally expect such problems to require more information to be specified at the free boundary than would be necessary for well-posedness of the classical initial value or boundary value problems that we have considered hitherto. Deciding how much extra information is necessary is one of the first challenges for the development of the theory; in many examples the modelling process suggests the correct information to be prescribed, but this is not always the case.

The principal reason why this topic has only recently gelled into a coherent subject can be traced to the spread of applied mathematics into areas of science which previously lacked detailed quantitative analysis. These applications revealed many new examples of free boundary problems and we begin by listing some of them. In most cases we will be dealing with 'field' equations identical to those

[132]The equation of curve shortening (6.88) may be used to describe the evolution of a curve, but we will not refer to this as a free boundary problem.

$$\left[k\nabla u \cdot \mathbf{n}\right]_{\text{solid}}^{\text{liquid}} = -\rho L v_n. \tag{7.4}$$

Note that, if we write Γ as $f(x, y, z, t) = 0$, then $\mathbf{n} = \nabla f / |\nabla f|$ and

$$v_n = -\frac{\partial f}{\partial t} \Big/ |\nabla f|. \tag{7.5}$$

The condition (7.4) is the 'extra' information referred to above; if $L = 0$ and there is no phase change, we can simply read off Γ as the isotherm $u = u_m$.

This model is very simplistic and we need to make several remarks. First, we have made no comment about whether or not $u > u_m$ in the liquid or $u < u_m$ in the solid. We will see later that the question of whether these inequalities hold or not has far-reaching implications for the mathematical structure of the model. Second, the model is a gross simplification of what happens in most practical melting or freezing situations, because we have neglected many important phenomena such as convection, radiation and, most of all, the presence of impurities. The latter is one of the reasons why u_m should not necessarily be thought of as a prescribed constant, and it can be shown that even small variations in u_m can drastically alter the predictions.[134] Nonetheless, the Stefan problem is challenging mathematically and at once reveals all the nonlinearity inherent in any free boundary problem. Even though the field equation (7.1) is linear, we can never superimpose two solutions of a free boundary problem and hope to generate a third because their domains of definition are different, and this neutralises many of the techniques we have been describing in previous chapters. In fact, the only obvious simplification we are ever likely to be able to make to a Stefan problem is to consider its 'one-phase' specialisation. If, say, the ice is at $u \equiv u_m - 0$ or the water at $u \equiv u_m + 0$, then clearly we need only solve on one side of Γ, which may ease our computational task. Any free boundary problem in which the solution is trivial on one side of the free boundary will henceforth be referred to as *one-phase*.

There are many, many models related to the Stefan problem.[135] Suppose, for example, we neglect the specific heat term, i.e. the left-hand side of (7.1), still of course retaining the time derivative in (7.5), and also consider a one-phase problem. This leaves us with the *Hele-Shaw* free boundary problem in the theory of viscous flow. The basic idea of its derivation there is as follows. Viscous fluid is forced either by pumping or suction through the narrow gap between two parallel plates $z = 0$ and $z = h$, as in Fig. 7.2. The equations relating the two 'in-plane' velocity components (u, v) to the pressure p are quite simple when we neglect inertia and the in-plane shear force, and only allow for the shear that occurs as we traverse the gap from $z = 0$ to $z = h$. We find

$$\mu\frac{\partial^2 u}{\partial z^2} = \frac{\partial p}{\partial x}, \quad \mu\frac{\partial^2 v}{\partial z^2} = \frac{\partial p}{\partial y},$$

[134] This is as we might hope, since only a small change in composition can change iron into steel, or turn an icy road into a wet one, by the addition of carbon or salt, respectively.

[135] An interesting example is that of Type I superconductivity, where, if the geometry is right, the scalar magnetic field H satisfies the parabolic form of Maxwell's equations (as in (6.11)) and the 'Meissner effect' requires $H = 0$ in the superconducting region. The Stefan condition follows from a balance of magnetic flux.

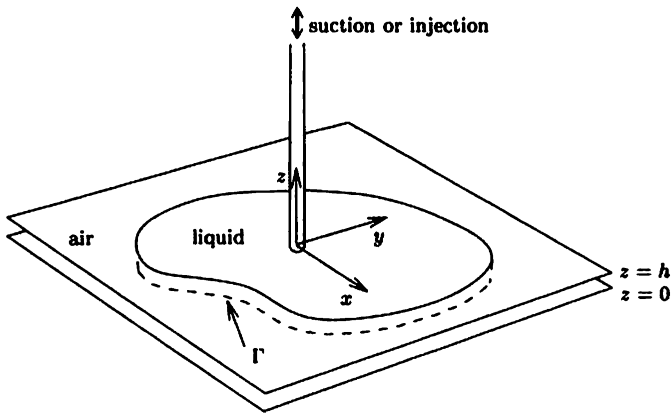


Fig. 7.2 A Hele-Shaw cell.

so that

$$\begin{pmatrix} u \\ v \end{pmatrix} = -\frac{z(h-z)}{2\mu}\nabla p, \tag{7.6}$$

where $\nabla = (\partial/\partial x, \partial/\partial y)$. Then, conservation of mass requires that the third velocity component w be related to u and v by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{7.7}$$

so, integrating (7.7) with respect to z and using the fact that $w = 0$ on $z = 0$ and $z = h$, (7.6) gives

$$\nabla^2 p = 0. \tag{7.8}$$

Finally, suppose that when we look down on the (x, y) plane of the cell, we see a free boundary Γ, as in Fig. 7.2. Conservation of momentum can be used to assert that the pressure satisfies

$$p = \text{constant} \tag{7.9}$$

along Γ, which is analogous to (7.3), and conservation of mass requires that the normal velocity of Γ should equal the average of $(u, v) \cdot \mathbf{n}$ across the gap between the plates.[136]

Hence

$$-\frac{\partial p}{\partial n} = \text{constant} \cdot v_n, \tag{7.10}$$

which is the one-phase version of (7.4). It is a simple exercise using (7.5) to show that (7.9) and (7.10) can be written as

[136]We will see these conditions in another fluid-dynamical context shortly. Their validity for the Hele-Shaw cell is contentious from the modelling viewpoint.

$$\frac{\partial p}{\partial t} - |\nabla p|^2 = 0, \tag{7.11}$$

with a suitable rescaling of p.

One of the great virtues of this model is that it has a simple generalisation which is one of the easiest free boundary problems to visualise: simply squeeze together two transparent sheets with a drop of viscous liquid, such as shampoo or treacle, sandwiched between them, because setting $w = \partial h/\partial t$ on $z = h$ merely converts (7.8) into a Poisson equation with a term proportional to $\partial h/\partial t$ on the right-hand side. However, the Hele-Shaw problem also models the important process of electrochemical machining, where a metal 'workpiece' is to be eaten away into a desired shape by being immersed in an electrolyte. Then, if ϕ is the electric potential in the relatively poorly conducting electrolyte, and we are away from the thin layers of high charge density near the workpiece, the argument in §5.1.2 gives

$$\nabla^2\phi = 0, \tag{7.12}$$

with, say, $\phi = 0$ on the workpiece. Moreover, it is observed that dissolution at the workpiece occurs at a rate proportional to the local electric field, so that, again,

$$\frac{\partial \phi}{\partial n} = \text{constant} \cdot v_n; \tag{7.13}$$

the sign of ϕ in the problem changes if deposition occurs, as in the process of electroforming.

Almost the same model arises in porous medium flow in the absence of gravity. Suppose the flow is modelled as in §5.1.4.4 but that, as in Fig. 7.3, the porous medium contains a free boundary Γ separating adjacent regions, in one of which the pores are full of water, so that the medium there is 'saturated', and in the other the pores are completely dry (partial saturation always occurs in reality, but often only over quite thin regions). We again have Laplace's equation for the pressure in the saturated region and the arguments leading to the Hele-Shaw free boundary conditions can be repeated. If, on the other hand, gravity is important, so that Darcy's law is

$$\mathbf{v} = -\nabla(p + y) \tag{7.14}$$

in non-dimensional variables with y vertical, then, in incompressible flow,

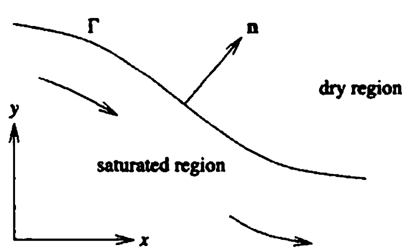


Fig. 7.3 Porous medium flow.

$$\nabla \cdot \mathbf{v} = -\nabla^2 p = 0. \tag{7.15}$$

The mass conservation condition at the free boundary $f = 0$, where p vanishes, becomes

$$\frac{\partial f}{\partial t} - \nabla p \cdot \nabla f = \frac{\partial f}{\partial y} \tag{7.16}$$

(see Fig. 7.3).[137] Since $p = 0$, we can, as in (7.11), write this condition as

$$\frac{\partial p}{\partial t} - |\nabla p|^2 = \frac{\partial p}{\partial y}.$$

A further generalisation would be to suppose that the free boundary separates two immiscible fluids with different viscosities $\mu_{1,2}$, such as oil and water. Then, in the absence of any interfacial forces, such as capillarity, the momentum and mass balances at Γ yield the so-called *Muskat problem*, whose free boundary conditions in suitable dimensional variables are

$$p_{\text{water}} = p_{\text{oil}}, \quad \mu_{\text{water}}^{-1}\frac{\partial p_{\text{water}}}{\partial n} = \mu_{\text{oil}}^{-1}\frac{\partial p_{\text{oil}}}{\partial n} = -v_n$$

in the absence of gravity. The Hele-Shaw model can then be retrieved as a one-phase limit (see Exercise 7.2).

7.1.2 Other free boundary problems in diffusion

An interesting problem in finance describes the valuation of a particular kind of option called an *American put option.* This differs from the 'European' call option described in §6.1.2 in two ways. The first is that the holder has the right to sell the asset for a specified amount E, instead of buying it. The second is that the right to sell can be exercised at *any* time up to the pre-assigned expiry date, $t = T$, rather than only at $t = T$. This introduces the idea of an 'optimal exercise price' into the modelling and converts what was a conventional parabolic equation model into a free boundary problem in which the free boundary is the graph of the optimal exercise price as a function of time t.

Suppose we have the flexibility inherent in American options. To accord with the absence of arbitrage, the option value must be at least equal to the proceeds of exercising it, since exercise can take place at any time. We can extend the discussion after (6.10) to note that, if $S < E$, then the net result of buying the option, at a cost of $-V$, and exercising it immediately to yield $E - S$, is $E - S - V$. Hence, in the absence of 'free lunches', for $0 \leqslant t \leqslant T$, $V \geqslant E - S$, and a similar argument shows that we must also have $V \geqslant 0$. Furthermore, if S falls to zero, then it remains there, and the holder should obviously exercise the option immediately in order to gain from the interest on the proceeds. There must therefore be a range of values $0 < S < S^*(t)$ in which the option should be exercised, so in this range

[137] The configuration in which the porous medium is a rectangle separating two reservoirs of water of different prescribed depths is now so famous that it is referred to as the *dam problem*.

its value is $E - S$, while for $S^*(t) < S < \infty$ the Black–Scholes equation still holds. We therefore seek solutions of

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = r\left(V - S\frac{\partial V}{\partial S}\right) \quad \text{for } S^*(t) < S < \infty, \tag{7.17}$$

where $V \to 0$ as $S \to \infty$ and, if the option is still unexercised at $t = T$,

$$V(S,T) = \max(E - S, 0). \tag{7.18}$$

As explained above, we also expect that

$$V(S,t) \geqslant \max(E - S, 0) \quad \text{for } 0 \leqslant t \leqslant T, \tag{7.19}$$

with equality only holding in the region $0 \leqslant S \leqslant S^*(t)$, where the option is exercised. This simple-minded argument does not give a free boundary problem for V in the style of §7.1.1. However, further consideration of the optimal exercise strategy (see Exercise 7.4) reveals that V satisfies the free boundary conditions

$$V = E - S, \quad \frac{\partial V}{\partial S} = -1 \quad \text{at } S = S^*(t), \tag{7.20}$$

these two free boundary conditions being analogous to specifying the melting temperature and balancing the energy at the free boundary of a Stefan problem.

The question of determining what extra information should be assigned at the free boundary is also illustrated by the modelling of flames as combustion problems, usually in gases, in which the important chemical reactions only occur in the flame sheet. In many cases of practical interest the thickness of flame sheets is much less than the typical dimensions of the environment in which the flame exists, and hence a free boundary model might be appropriate. Unfortunately, combustion modelling is such a complicated subject that we can only mention two simple free boundary approaches that can be used to predict flames. The relevance of each approach depends crucially on whether the reactants are provided 'ready-mixed' and capable of being ignited, say by a spark, or whether diffusional mass transfer has to occur in order to create the correct 'stoichiometric' mixture.

The former case leads to what is called a *premixed flame*, and in theory it should be modelled by coupled equations for the temperature T and reactant concentration c, as in §6.7.2. The key modelling assumption that leads to a free boundary problem is that the dimensionless rate of reaction is, as in §§5.11.1 and 6.6.1, of the form $Ace^{-E/T}$, where A is constant and the dimensionless 'activation energy' E is large. This implies that the reaction term is negligible except where the temperature is near its highest value, say T_0, which is now an unknown of the model. It can then be expected that the reaction is confined to a thin region where T is close to T_0 but where c is still appreciable. Assuming purely conductive heat transfer (as, say, for combustion in a solid such as a cigarette), in one dimension we are led to the model

$$\frac{\partial T}{\partial t} = k\frac{\partial^2 T}{\partial x^2} \tag{7.21}$$

in the region ahead of the flame, say $x > s(t)$, with

$$T = T_0 \tag{7.22}$$

at and behind the flame, $x \leqslant s(t)$. We now see another one-phase free boundary problem beginning to emerge in $x > s(t)$, but, at the moment, we know neither T_0 nor $\partial T/\partial x$ as $x \downarrow s(t)$.

We can make some further progress by assuming that the mass consumption is proportional to the rate of reaction, so that the reaction zone near the flame is modelled by

$$\frac{\partial T}{\partial t} = k\frac{\partial^2 T}{\partial x^2} + Ace^{-E/T}, \quad \frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - \alpha ce^{-E/T}, \tag{7.23}$$

where α is a constant analogous to A. As discussed in §6.7.2, such parabolic systems are not easy and here we make the further assumption that the reaction front is a *slow* travelling wave solution of (7.23) in which T is close to T_0. Thus, when we neglect $\partial/\partial t$ and notice that

$$c \to 0, \quad T \to T_0 \tag{7.24}$$

behind the flame, we can add and integrate to find that

$$\alpha kT + ADc = \alpha kT_0. \tag{7.25}$$

Finally, approximating $e^{-E/T}$ by $e^{-E/T_0}e^{E(T-T_0)/T_0^2}$, as in §6.6.1, we can integrate

$$\frac{\partial^2 T}{\partial x^2} + \frac{\alpha}{D}e^{-E/T_0}(T_0 - T)e^{E(T-T_0)/T_0^2} = 0$$

once, and use (7.24) again and the fact that $T - T_0$ is relatively large and negative ahead of the flame, to give

$$-\left.\frac{\partial T}{\partial x}\right|_{x=s(t)} = \sqrt{\frac{2\alpha}{DE}}T_0 e^{-E/2T_0}. \tag{7.26}$$

This gives the second Stefan condition to go with (7.22), but T_0 is still undetermined; we will return to this question in §7.5.1.1.

A quite different situation occurs when the ambient combustible material is not supplied in a premixed form. Suppose instead that there are just two reacting components, for example fuel at concentration c_1 and oxygen at concentration c_2, and that they only react where they diffuse into each other. For 'fast' reactions, the concentration of oxygen is negligible on the 'fuel side' of the flame, and vice versa, in which case we simply have to solve diffusion equations for c_i on either side of the flame, at which $c_1 = c_2 = 0$. Also, since all the chemical reaction takes place at the free boundary, a 'stoichiometric' condition holds there. This is jargon for the fact that, because chemicals react in fixed ratios (for example, two hydrogen molecules and one oxygen molecule combine to create water), the rates at which chemicals are used up at a flame are proportional to each other. Hence there is a linear relationship between the two mass flows,

$$\frac{\partial c_1}{\partial n} = -\lambda \frac{\partial c_2}{\partial n},$$

where λ is a constant. This is a model for so-called *diffusion flames*, and contrasts with the Muskat problem.

Flame modelling in gases is often further complicated by the presence of convection, which is strongly coupled to the equations of heat and mass diffusion, and can lead to phenomena such as quenching or, with compressibility important, to detonation or deflagration, as discussed in §2.5.3.

Less dramatic diffusion processes may lead to free boundary problems which are so 'mild' that the free boundary conditions may be implicit in the modelling statement. In fact, we have already encountered such a situation with models leading to the porous medium equation (6.72), $\partial p/\partial t = \nabla \cdot (p^n \nabla p)$, in the derivation of which we tacitly assumed that the pressure p (or the film thickness h if that is the dependent variable) is positive. However, it is easy to imagine that the 'active' region in which $p > 0$ abuts an inert one in which $p \equiv 0$. Now $p \equiv 0$ is a solution of (6.72), so this prompts the question 'suppose we have an initial value problem in which the data vanishes in some region and is strictly positive elsewhere; does it make sense to try to solve the equation as it stands for $t > 0$ and hope the free boundary separating the active and inert regions comes out in the wash?' Strictly speaking, the mathematical answer to this question is no, unless the derivatives of p are such as to make $\nabla \cdot (p^n \nabla p)$ vanish at the free boundary; only in such a case could the left- and right-hand sides of (6.72) be equal there. Everything boils down to the smoothness of the solution and we will have to focus on this in our later discussion of the porous medium equation.

Another interesting situation occurs in the *oxygen consumption problem* in which oxygen is removed from a biological tissue by chemical action at a rate which is approximately independent of the amount of oxygen remaining, just as long as the latter is positive. A very simple one-dimensional model for the oxygen concentration c is

$$\frac{\partial c}{\partial t} - \frac{\partial^2 c}{\partial x^2} = -1, \tag{7.27}$$

which at first sight seems to be a simple linear case of the reaction–diffusion problems studied in §6.6. However, the simple solution $c(x,t) = \text{constant} - t$ shows that we must take care to enforce the physical requirement that $c \geqslant 0$. Thus, in line with some earlier models in this chapter, we could propose a one-phase free boundary model in which regions where $c > 0$ adjoin regions where $c \equiv 0$, and (7.27) only holds in the former. The two free boundary conditions would be $c = 0$ and, by conservation of mass, $\partial c/\partial x = 0$. However, an alternative approach, in the spirit of Chapter 6, would be to replace the right-hand side of (7.27) by $-H(c)$, where

$$H(c) = \begin{cases} 1, & c > 0, \\ 0, & c \leqslant 0, \end{cases}$$

the Heaviside function, is clearly a nonlinear function of c.[138] The relationship between these approaches thus leads to yet another mathematical question akin to those raised in connection with the porous medium equation and with the model for American options, and we will see that its answer lies in some interesting theories of weak solutions for parabolic equations.

7.1.3 Some other problems from mechanics

In Chapter 2 we have already encountered one of the most intensively studied free boundary problems, namely that of shock waves in compressible fluids. We recall that the most difficult aspect of our discussion there concerned the question of uniqueness, for either the classical formulation or the weak formulation. In both cases we found that, in order to obtain a unique physically acceptable solution, we needed to append an extra restriction to the Rankine–Hugoniot relations. This was despite the fact that these relations apparently already contain an extra equation over and above those that would be needed for a Cauchy problem with a smooth solution on a prescribed domain. For example, for the simple model

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0, \tag{7.28}$$

with Cauchy data $u(x,0) = u_0(x)$, a count of the number of pieces of information would not lead us to expect to be given any data other than the Rankine–Hugoniot condition in order to be able to determine the unknown position, $x = s(t)$, of a shock. However, our discussion at the end of §1.7 tells us that the Rankine–Hugoniot condition allows us too much freedom and that a further restriction is needed to obtain a unique physically acceptable solution.

Bearing this idea in mind, we now list some further free boundary problems in mechanics.

7.1.3.1 Fluid dynamics

A venerable source of free boundary problems is the mechanics of continua that have a 'free surface', which in mechanics often means a 'stress-free' surface. The simplest case occurs in irrotational inviscid fluid dynamics which provides a model for, say, water in motion beneath an atmosphere which exerts no stress other than a constant pressure. Then, in the absence of surface tension, a momentum balance demands that the water pressure p is equal to atmospheric pressure. Now, as will be shown in Chapter 9, it is a simple deduction from the time-dependent generalisation of (2.5) and (2.6) to show that, when the density ρ is constant, p is given everywhere in the water by *Bernoulli's equation*

$$\frac{p}{\rho} + \frac{1}{2}|\nabla\phi|^2 + gy + \frac{\partial\phi}{\partial t} = \text{constant}, \tag{7.29}$$

where y is the vertical coordinate and ϕ is the velocity potential, which satisfies Laplace's equation. Hence, if we denote the water surface by $y = \eta(x,t)$, then, in a two-dimensional problem,

[138] This is a special case of the problem of 'dead cores' in more general chemical reactions in which the right-hand side of (7.27) is proportional to c^p, as mentioned at the end of §5.11.3.

$$\frac{1}{2}|\nabla\phi|^2 + g\eta + \frac{\partial\phi}{\partial t} = \text{constant} \quad \text{on } y = \eta. \tag{7.30}$$

Also, we know that, whether or not $y = \eta$ is stress free, no fluid particles can cross this surface, i.e.

$$\frac{\partial\phi}{\partial n} = v_n, \tag{7.31}$$

where, as usual, v_n is the normal velocity to the surface. Using (7.5), we have our second free boundary condition,

$$\frac{\partial\phi}{\partial y} = \frac{\partial\eta}{\partial t} + \frac{\partial\phi}{\partial x}\frac{\partial\eta}{\partial x} \quad \text{on } y = \eta. \tag{7.32}$$

It is possible to non-dimensionalise the variables, so that (7.30) becomes

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + \frac{1}{F}\eta = \text{constant}, \tag{7.33}$$

where F is a dimensionless number called the *Froude number*; the larger the value of F, the larger is the effect of fluid inertia relative to gravity.

We can already see an opportunity for 'technology transfer' from gas dynamics. For, suppose we replace the atmosphere by a second active inviscid irrotational fluid, with a constant density $\rho' \neq \rho$, could we not regard the free surface as a 'shock' and derive (7.30)–(7.32) directly? After all, the fluid dynamics equations were originally conservation statements, although we have rather lost sight of this when we transformed to the variable ϕ. Anyway, the answer to our question is in the affirmative because the Rankine–Hugoniot condition for the mass conservation equation, namely

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0, \tag{7.34}$$

is $[\rho]\,v_n = [\rho\mathbf{u}\cdot\mathbf{n}]$, which leads to (7.32) when $\rho' = 0$ and $\mathbf{u} = \nabla\phi$.[139]

The simplest class of such inviscid free surface flows and, indeed, the only class that is at all tractable as far as explicit solutions are concerned, occurs when

[139]For the momentum conservation equation in three dimensions, we must regrettably write the conservation law in suffix notation as

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(p\delta_{ij} + \rho u_i u_j) = 0, \tag{7.35}$$

where $\mathbf{u} = (u_i)$ and δ_{ij} is the identity matrix. We then find that

$$[p\mathbf{n} + \rho(\mathbf{u}\cdot\mathbf{n})\mathbf{u}] = [\rho\mathbf{u}]v_n,$$

which leads to (7.33) when $g = 0$. Note also that, when $\rho \neq \rho'$ and neither is zero, then (7.32) and (7.35) also admit the possibility that

$$[\mathbf{u}\cdot\mathbf{n}] = 0, \quad [p] = 0, \quad \mathbf{u}\cdot\mathbf{n} = v_n,$$

which is what we called a contact discontinuity on p. 59, and which is illustrated by (7.37).

the flow is steady and two-dimensional, and F is infinite in (7.33). Then the free boundary conditions for Laplace's equation become

$$\frac{\partial \phi}{\partial n} = 0, \quad |\nabla \phi| = 1, \tag{7.36}$$

without loss of generality, and we have what is called a *Helmholtz flow*.[140]

An interesting complication is introduced into these kinds of inviscid flows when, in a two-phase problem, we permit relative tangential motion to occur across the free boundary, even across a free boundary where there is no density discontinuity. Then we have what is called a *vortex sheet*, at which the free boundary conditions for the velocity potential ϕ_i on either side are

$$\frac{1}{2}|\nabla \phi_1|^2 + \frac{\partial \phi_1}{\partial t} = \frac{1}{2}|\nabla \phi_2|^2 + \frac{\partial \phi_2}{\partial t} + \text{constant}, \tag{7.37}$$

which is conservation of normal momentum, and

$$\frac{\partial \phi_1}{\partial n} = \frac{\partial \phi_2}{\partial n} = v_n, \tag{7.38}$$

which is conservation of mass.

Further possibilities arise when we consider *rotational* inviscid flows. In steady two-dimensional situations we have seen in §5.11.1 that the stream function $\psi(x, y)$ satisfies

$$-\nabla^2 \psi = f(\psi), \tag{7.39}$$

where $f(\psi)$ is the vorticity and is usually an unknown function. However, there is often experimental evidence that the vorticity is localised in a patch and, if this is the case, we could take f to be a function with compact support. The question then arises as to what free boundary conditions should be imposed at the boundary of the patch.[141] As for free surface flows, no particles can cross the free boundary Γ, which is thus a level curve of ψ in steady flow, with ψ continuous at Γ. Also, since the pressure must be continuous to ensure momentum conservation, the Bernoulli condition requires that the tangential velocity $\partial \psi / \partial n$ must be continuous at Γ.[142]

The list of problems of this type is long enough, even without introducing the effects of viscosity, which leads to considerable algebraic complications (see Exercise 7.6). Hence our final examples come from solid mechanics.

7.1.3.2 Solid mechanics

Stress-free boundaries are also ubiquitous in solid mechanics, but fortunately many elastic bodies only undergo such small displacements that their free surfaces may be regarded as being more or less in prescribed positions. This enables previously

[140] These free boundary conditions also arise in certain very simple models of flame propagation [7].

[141] Precisely the same problem arises when we model the magnetic confinement of a plasma (an ionised gas) in a long tube.

[142] Again, we will have to wait until Chapter 9 to explain the Bernoulli condition.

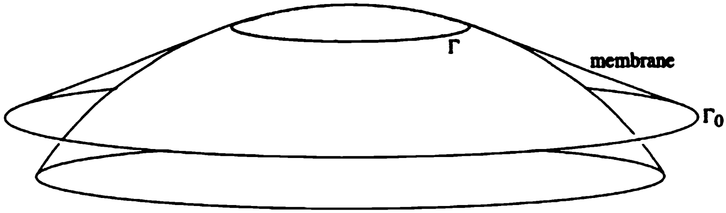


Fig. 7.4 Obstacle problem for a membrane.

mentioned theories of linear elasticity to be used, but there is one situation where nonlinearity cannot be avoided, even though the displacements are small. This happens whenever two elastic bodies come into contact because, even though the small transverse displacements on a smooth contact region can be regarded as part of the data for the problem, the perimeter of the contact region is an unknown free boundary. We have already encountered such problems in §§5.1.6.2 and 5.9.5, but, because we only considered two-dimensional configurations there, the contact perimeter consisted of discrete points. We showed that their location could be determined by applying appropriate singularity conditions in their vicinity and equations such as (5.151) reveal the nonlinearity of the problem. Similar contact problems in three-dimensional elasticity reveal the same dimension deficit, in that the lowest-order problem is to find the one-dimensional perimeter of a two-dimensional contact set. We will return to this topic of 'codimension-two' free boundaries at the end of the chapter, but there is one special case that is relatively easy to model yet leads to a conventional 'codimension-one' free boundary problem. This is when one of the bodies is rigid and smooth[143] and the other thin and extensible enough to be modelled as a smooth membrane, as in Fig. 7.4.

We then have what is called an *obstacle problem*, which is simple to state if we assume that the membrane is stretched, with a fixed perimeter Γ_0 and only has a small deflection when it comes into contact with the rigid obstacle $z = f(x, y)$. As in §5.1.4.2, in equilibrium the transverse displacement $u(x, y)$ satisfies

$$\nabla^2 u = 0 \tag{7.40}$$

wherever the membrane is not in contact, i.e. when $u > f$, but otherwise $u = f$ and we have a free boundary Γ to the contact region. Although it is obvious that $u = f$ on Γ, we again have our familiar question of what extra information needs to be imposed there. This situation is simple enough that we can see at once that equilibrium can only be maintained in the vicinity of Γ if the tension T is continuous and the force $T\nabla u \cdot \mathbf{n}$ is the same as we approach Γ from either side; this is the same as carrying out the balance that led to (7.30), with the free boundary being a curve rather than a surface. Hence our free boundary conditions are simply

[143] If the contacting bodies are rough, then the interesting problem arises of determining another free boundary, namely that between the regions of sliding contact and the regions of adhesion.

$$u = f \tag{7.41}$$

and

$$\mathbf{n} \cdot \nabla u = \mathbf{n} \cdot \nabla f. \tag{7.42}$$

These conditions are analogous to the singularity conditions that we imposed on the contact or crack-closing problems described in Chapter 5. Other free boundary (but not stress-free boundary) problems arise in elastic-plastic deformation, where the boundary of the region of plastic flow is to be determined, and in various other solid mechanics problems involving fracture or slip.

We now continue with our standard procedure of trying to describe the mathematical ideas with which the above problems can be analysed and, sometimes, solved. We begin by discussing well-posedness, which provides some surprises when compared to our discussion of this topic in Chapters 4–6.

7.2 Stability and well-posedness

As already mentioned, the inevitable nonlinearity of free boundary problems makes them less susceptible to mathematical analysis than the linear equations on which we have largely focused in the last three chapters. The only general approach that is available is to study the response of the system in the vicinity of a solution, be it ever so trivial, that we are lucky enough to know explicitly. In fact, we adopted this philosophy when we discussed qualitative properties of general second-order equations; in Chapters 2 and 3, we sometimes 'froze' the coefficients of a nonlinear equation at prescribed constant values and looked for the local behaviour, as described by the resulting linear equation. We hope that such an investigation can at least tell us how robust the model is to small perturbations, and now we apply the idea to some free boundary problems.

Before we start, we remark that the systematic application of small perturbation theory in applied mathematics often relies on the ideas of asymptotic expansion, which we have deliberately eschewed in this book. For the remaining chapters we make one concession to asymptotic analysis by using the notation $A \sim B$ as $\epsilon \to 0$ to express the fact that $A/B \to 1$ as $\epsilon \to 0$. As usual, we quantify the difference $A - B$, for example, as $O(\epsilon)$ ('roughly the same size as ϵ') or $o(\epsilon)$ ('much smaller than ϵ'); see [22] for precise details.

Our procedure, which often goes under the heading of *linear stability theory*, follows ideas introduced in §5.11.3 and taken further in §6.6.4. We suppose we know an explicit solution of the field equation(s) which is compatible with the free boundary conditions, for simplicity restricting ourselves to one-dimensional problems. We then seek a solution in which the free boundary has an asymptotic expansion

$$x = s(t;\epsilon) \sim s_0(t) + \epsilon s_1(t) + \cdots \tag{7.43}$$

in terms of some small parameter ϵ which measures the size of the disturbance that we are considering. Since the free boundary conditions are to be evaluated at $x = s(t,\epsilon)$, they can also be expanded asymptotically; typically, we can write the value of any function $f(x,t)$ on the free boundary as

$$f(s(t;\epsilon),t) \sim f(s_0(t),t) + \epsilon s_1(t)\frac{\partial f}{\partial x}(s_0(t),t) + \cdots . \tag{7.44}$$

Hence, when we write the solution of the field equations as

$$u(x,t;\epsilon) \sim u_0(x,t) + \epsilon u_1(x,t) + \cdots ,$$

we find that

$$u(s(t;\epsilon),t;\epsilon) \sim u_0(s_0(t),t) + \epsilon\left(u_1(s_0(t),t) + s_1(t)\frac{\partial u_0}{\partial x}(s_0(t),t)\right) + O(\epsilon^2). \tag{7.45}$$

It is important always to remember to include both the terms in the large brackets in (7.45).

Now u_0 is known and hence, assuming we are allowed to equate terms of $O(\epsilon)$ to zero independently, we are left with boundary conditions that are not only linear in u_1, but are to be evaluated on the known boundary $s_0(t)$. Hence the problem for u_1 should be amenable to the methods of the previous three chapters. Unfortunately, it is rare for us to know enough about the original free boundary problem to be able either to justify the key step of equating the coefficient of ϵ to zero or to assess the quality of the approximation $u_0 + \epsilon u_1$ to the full solution rigorously. However, we will be undeterred by this fact since we will see from the examples below that predictions from this linearised approach agree so well with many pieces of physical evidence. Surprisingly, when we start to consider various problems for u_1, we will find that free boundary problems are prone to instability and ill-posedness far more than any of the models considered in previous chapters. The interesting examples are more than one-dimensional, so we will need to generalise (7.45) to show that the value of a function $u(x,y,t;\epsilon)$ on a free boundary $x = s(y,t;\epsilon) \sim s_0(t) + \epsilon s_1(y,t) + \cdots$ is given by

$$\begin{aligned} u(s(y,t;\epsilon),y,t;\epsilon) &\sim u_0(s_0(t),y,t) \\ &\quad + \epsilon\left(u_1(s_0(t),y,t) + s_1(y,t)\frac{\partial u_0}{\partial x}(s_0(t),y,t)\right) + O(\epsilon^2). \end{aligned}$$

(In the next example we will follow convention in that the unperturbed interface is $y = 0$, but this should cause no confusion.)

7.2.1 Surface gravity waves

A very well-documented example concerns the solution of Laplace's equation for a velocity potential $\phi(x,y,t)$ in $y < \eta(x,t)$, subject to (7.32) and (7.33), in the vicinity of the zero solution $\phi = 0$, $\eta = 0$ (the constant in (7.33) can be taken to be zero without loss of generality). To aid the bookkeeping, we introduce a small parameter, as in (7.43), and write

$$\phi = \epsilon\phi_1, \quad \eta = \epsilon\eta_1,$$

to give

$$\nabla^2\phi_1 = 0 \quad \text{in } y < \epsilon\eta_1, \tag{7.46}$$

with

$$\epsilon\left(\frac{\partial\phi_1}{\partial y} - \frac{\partial\eta_1}{\partial t}\right) + O(\epsilon^2) = 0 \tag{7.47}$$

and

$$\epsilon\left(\frac{\partial\phi_1}{\partial t} + \frac{1}{F}\eta_1\right) + O(\epsilon^2) = 0 \tag{7.48}$$

on $y = \epsilon\eta_1$. Approximating as above, we replace these free boundary conditions by

$$\frac{\partial\phi_1}{\partial y} = \frac{\partial\eta_1}{\partial t}, \quad \frac{\partial\phi_1}{\partial t} + \frac{1}{F}\eta_1 = 0 \quad \text{on } y = 0. \tag{7.49}$$

Cross-differentiation gives

$$\frac{\partial^2\phi_1}{\partial t^2} + \frac{1}{F}\frac{\partial\phi_1}{\partial y} = 0 \quad \text{on } y = 0, \tag{7.50}$$

and we are left with an unconventional boundary value problem for the potential ϕ_1 in $y < 0$. The presence of the second time derivative means that it does not fall into any of the categories in Chapters 4–6; however, since the coefficient of this time derivative is constant in time, which always happens when the unperturbed solution is independent of time, we can seek solutions as Fourier superpositions of $\cos\lambda t$ and $\sin\lambda t$, where $\lambda =$ constant. It makes life easier if we write

$$\phi_1(x,y,t) = \Re\left(e^{i\lambda t}\tilde{\phi}_1(x,y)\right)$$

and then seek values of λ that yield non-trivial solutions for $\tilde{\phi}_1$. The fact that (7.46) and (7.50) have no forcing terms means that we are led to an eigenvalue problem for λ, and we will be tempted to regard the problem as (linearly) ill-posed if and only if at least some of these eigenvalues have negative imaginary part. In such a case, an arbitrary small spatial disturbance grows exponentially rapidly in time; equally, if all eigenvalues have positive imaginary part, then we regard the problem as linearly stable, and if there are only eigenvalues with zero and positive imaginary part it is *neutrally stable*. Of course, to obtain an eigenvalue problem we must first close the model for ϕ. Hence we need other boundary conditions that describe how the liquid is contained; for simplicity, we assume that it extends to $y = -\infty$ and that $-\infty < x < \infty$, and that it is at rest at large distances from the surface.

Carrying out this procedure, we obtain

$$\nabla^2\tilde{\phi}_1 = 0, \tag{7.51}$$

with the Robin condition

$$-\lambda^2 F\tilde{\phi}_1 + \frac{\partial\tilde{\phi}_1}{\partial y} = 0 \quad \text{on } y = 0, \tag{7.52}$$

and the boundedness condition

$$\tilde{\phi}_1 \to 0 \quad \text{as } y \to -\infty.$$

We recall from §5.2.1 that the sign in the Robin condition is vital in deciding the uniqueness of solutions of Laplace's equation, and here, if λ is real, this sign is such that $\tilde{\phi}_1 = 0$ is probably not the only solution to (7.51) and (7.52). By separation of the variables, we soon find that there are indeed non-zero solutions in which

$$\tilde{\phi}_1 = \Re\left(A(k)e^{ky \pm ikx}\right),$$

where $A(k)$ is an arbitrary function of k, just as long as $\Re k > 0$ and

$$\lambda = \pm\sqrt{\frac{k}{F}}. \tag{7.53}$$

Physically, this corresponds to a 'wave train' on the free boundary in which η is proportional to $\Re e^{i(\pm kx + \lambda t)}$, so that the wave speed is λ/k, assuming this is real. Notice that $|\nabla\tilde{\phi}_1|$ is only bounded as $|x| \to \infty$ if k is real, and so we only have physically acceptable 'eigensolutions' of the linearised water wave problem if the eigenvalue λ is real, which corresponds to what we have called a continuous spectrum in Chapter 5. The fact that λ^2 must be positive reinforces our earlier statement about uniqueness.

Thus the rest state of water with a horizontal free surface is neutrally stable. The result (7.53) is often called a ***dispersion relation*** because it relates the wavelength of the disturbance, $2\pi/k$, to its temporal behaviour, described by λ. We can generate solutions of the Cauchy problem, i.e. the initial value problem for (7.46) and (7.50), by superimposing terms of the form $A(k)e^{\pm it\sqrt{k/F}}e^{ikx}e^{ky}$, perhaps even with complex values of k, which simply corresponds to taking a Fourier transform in x. Note that we must be careful about the initial conditions to be imposed on such Cauchy problems. For free surface flows we need to prescribe both the initial potential and surface elevation because of the appearance of two time derivatives in the free boundary conditions.

7.2.2 Vortex sheets

The approach of the previous section can be applied to many inviscid fluid free boundary flows, to consider capillary waves, waves on flowing fluids, or the effect of finite depth. Here we give just one other application to the stability of a vortex sheet separating two infinite fluids in $y > 0$, $y < 0$, with the same density and in the absence of gravity, flowing parallel to each other with velocity $(U_\pm, 0)$. We now have to solve for two potentials and the free surface,

$$\phi^+ \sim U_+x + \epsilon\phi_1^+ + \cdots, \quad \phi^- \sim U_-x + \epsilon\phi_1^- + \cdots, \quad \eta \sim \epsilon\eta_1 + \cdots,$$

where, from (7.38) and (7.37), the linearised free boundary conditions are

$$\frac{\partial\phi_1^+}{\partial y} = \frac{\partial\eta_1}{\partial t} + U_+\frac{\partial\eta_1}{\partial x}, \qquad \frac{\partial\phi_1^-}{\partial y} = \frac{\partial\eta_1}{\partial t} + U_-\frac{\partial\eta_1}{\partial x},$$

$$\frac{\partial\phi_1^+}{\partial t} + U_+\frac{\partial\phi_1^+}{\partial x} = \frac{\partial\phi_1^-}{\partial t} + U_-\frac{\partial\phi_1^-}{\partial x}$$

on $y = 0$. Writing

$$\phi_1^+ = \Re A_+(k)e^{i(\lambda t+kx)}e^{-ky}, \quad \phi_1^- = \Re A_-(k)e^{i(\lambda t+kx)}e^{ky},$$

where $\Re k > 0$, and assuming $U_+ \neq U_-$, we soon find the dispersion relation

$$\lambda = \frac{k}{2}\left(-(U_+ + U_-) \pm i(U_+ - U_-)\right).$$

A similar result applies when the phase of ϕ is $\lambda t - kx$; we find

$$\lambda = \frac{k}{2}\left((U_+ + U_-) \mp i(U_+ - U_-)\right).$$

Hence, if we require k to be real so that $|\nabla\phi^\pm|$ are bounded at infinity, any wave train on the vortex sheet *grows* exponentially in time whenever $U_+ \neq U_-$. The situation is almost as bad as for the (ill-posed) backward heat equation; since $|\lambda| = O(k)$ as $k \to \infty$, the larger k is (i.e. the shorter the wavelength), the greater is the temporal growth rate, although it is not as rapid as the $O(k^2)$ growth for the backward heat equation. This is our first encounter with an everyday free boundary problem which appears to exhibit ill-posedness when we try to predict its evolution forward in time. Its catastrophic consequences can be seen the moment we consider solving an initial value problem for $\phi_1^\pm$ and η_1, with initial surface displacement $\eta_1(x,0) = \eta_{10}(x)$ by, say, taking a Fourier transform in x. The resulting solution, say for η_1, contains terms such as

$$\int_{-\infty}^{\infty} \hat{\eta}_{10}(k)e^{-ikx}e^{i\lambda(k)t}\,dk,$$

and, even if η_{10} is so well-behaved that $\hat{\eta}_{10} = O(\exp(-\alpha|k|))$ as $|k| \to \infty$, this integral diverges as soon as $t \geqslant 2\alpha/|U_+ - U_-|$. In other words, finite-time blow-up can occur for arbitrarily small time, even with very smooth initial data, as was the case for the backward heat equation in §6.4.2. In order to emphasise how frequently this situation can occur, we now describe the linear stability analysis of one other prototypical example.

7.2.3 Hele-Shaw flow

For the one-phase Hele-Shaw problem described in §7.1.1, we can easily see that there is an exact travelling wave solution in which the free boundary is again straight. When we take the free boundary conditions as $p = 0$ and $\partial p/\partial n = -v_n$ for simplicity, it is

$$p(x,y,t) = -V(x - Vt) \quad \text{for } x < Vt,$$

as long as the pressure gradient is V. Hence $V < 0$ represents a free boundary that retreats into the liquid under suction, while $V > 0$ represents an advancing boundary. Our usual linearisation procedure gives

$$p(x,y,t) \sim -V(x - Vt) + \epsilon p_1(x,y,t) + \cdots,$$

with free boundary

$$x \sim Vt + \epsilon\eta_1(y,t) + \cdots,$$

leading to

$$\nabla^2 p_1 = 0 \quad \text{for } x < Vt,$$

and the linearised boundary conditions

$$\epsilon\,(p_1 - V\eta_1) + O(\epsilon^2) = 0$$

and

$$\begin{aligned} V + \epsilon\frac{\partial\eta_1}{\partial t} + O(\epsilon^2) &= -\mathbf{n}\cdot\nabla p \\ &= -\left(1, -\epsilon\frac{\partial\eta_1}{\partial y}\right)\cdot\left(-V + \epsilon\frac{\partial p_1}{\partial x}, \epsilon\frac{\partial p_1}{\partial y}\right) + O(\epsilon^2) = V - \epsilon\frac{\partial p_1}{\partial x} + O(\epsilon^2) \end{aligned}$$

on $x = Vt + \epsilon\eta_1 + O(\epsilon^2)$. Hence, to lowest order,

$$p_1 = V\eta_1, \quad \frac{\partial p_1}{\partial x} = -\frac{\partial\eta_1}{\partial t} \quad \text{on } x = Vt.$$

Now the time t appears explicitly in the location of the boundary of our linearised problem, so we must define a travelling wave variable $\xi = x - Vt$ before we can seek a solution that is proportional to an exponential in t. This gives

$$\frac{\partial^2 p_1}{\partial\xi^2} + \frac{\partial^2 p_1}{\partial y^2} = 0 \quad \text{in } \xi < 0, \tag{7.54}$$

with

$$p_1 = V\eta_1, \quad \frac{\partial p_1}{\partial\xi} = -\frac{\partial\eta_1}{\partial t}$$

on $\xi = 0$. Hence, if we assume on physical grounds that $|\nabla p_1|$ is bounded as $\xi \to -\infty$, then we must have $p_1(\xi, y, t) = \Re A(k)e^{i(\lambda t \pm ky)}e^{k\xi}$, $k > 0$, and

$$i\lambda = -kV. \tag{7.55}$$

Hence, we have another dramatic switch from apparent well-posedness to ill-posedness as we go from a 'blowing' problem, $V > 0$, to a 'suction' problem, $V < 0$.

We remark that the Hele-Shaw analogy suggests that one-phase Stefan problems in which either supercooling (liquid temperature below u_m) or superheating (solid temperature above u_m) occurs are unstable; in either case the unstable boundary recedes into the active phase. Unfortunately, the algebraic complexity is too great for us to describe the two-dimensional stability analysis that justifies this statement.

These last two examples lead us to the realisation that, whereas all the partial differential equation models that we encountered in Chapters 4–6 were really predicated on the idea that well-posedness was a *sine qua non* for them to be accessible to mathematical study, with free boundary problems we have been led into a world where, if our stability arguments are a reliable guide, there is just about as much likelihood of ill-posedness as of well-posedness. In fact, many of the problems described in §7.1 simply switch their stability characteristics according to the direction of propagation of the free boundary or the sign of the driving mechanism. We can see this by changing the sign of F in (7.53), which corresponds to reversing gravity or upturning a glass of water, and the same thing often happens in diffusion problems.

7.2.4 Shock waves

Equally interesting, and perhaps even more far-reaching, is the outcome of a linear stability analysis of the simple hyperbolic equation (7.28),

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0. \tag{7.56}$$

We consider small one-dimensional perturbations to the travelling wave

$$u = \begin{cases} U_+, & x > Vt, \\ U_-, & x < Vt, \end{cases} \tag{7.57}$$

where the free boundary, i.e. the shock wave, has velocity V satisfying the Rankine–Hugoniot condition

$$V = \frac{1}{2}(U_+ + U_-). \tag{7.58}$$

Linearised perturbations to the field equation are found by writing $u \sim U_\pm \pm \epsilon u_1^\pm + \cdots$ in $x > Vt + \epsilon s_1(t) + \cdots$ and $x < Vt + \epsilon s_1(t) + \cdots$, respectively, to give, to lowest order,

$$\frac{\partial u_1^\pm}{\partial t} + U_\pm\frac{\partial u_1^\pm}{\partial x} = 0.$$

Hence

$$u_1^\pm(x,t) = \Re A_\pm e^{i\lambda_\pm(t - x/U_\pm)}, \quad A_\pm = \text{constant}.$$

Naive substitution shows that $u_1^\pm$ satisfy the perturbed Rankine–Hugoniot condition as long as

$$\frac{ds_1}{dt} = \frac{1}{2}\Re\left(A_- e^{i\lambda_- t(1-V/U_-)} + A_+ e^{i\lambda_+ t(1-V/U_+)}\right). \tag{7.59}$$

So it seems that *any* values of $\lambda_\pm$, real or complex, are permissible! That this is not in fact the case can only be seen by returning to the basic philosophy of our linear stability analysis, as follows.

Up to this point we have been content to write down solutions in which the free boundary suffered a small physically reasonable spatial perturbation, in particular a harmonic one, and seek whatever temporal behaviour would result from this spatial disturbance. In some cases we have gone further and asked about the implications for an initial value problem in which the initial data differed slightly

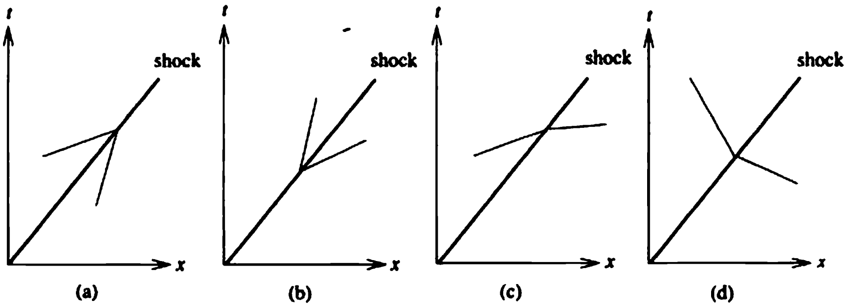


Fig. 7.5 Shock/characteristic orientation (cf. Figs 2.11 and 2.12).

from that which would have led to the unperturbed solution. In all these cases, this initial value problem has been solvable, in principle, by temporal and spatial superposition of the eigensolutions we have written down, although we have seen that blow-up may soon occur. All we needed was enough initial data to guarantee uniqueness, i.e. as many pieces of information as time derivatives in the model, in order to have the solution as a Fourier integral. However, when we adopt this approach for the hyperbolic equation (7.56), we see that we cannot even determine s_1 from (7.59) unless both the 'waves' $e^{i\lambda_\pm(t-x/U_\pm)}$ that propagate along the characteristics of the linearised problem impinge on the unperturbed free boundary $x = Vt$. In fact, (7.59) simply could not have been written down had we not made the implicit assumption that the relative configuration of the characteristics of (7.56) and the shock was as in Fig. 7.5(a) and not Fig. 7.5(b,c,d).[144] Put another way, we could only accept Fig. 7.5(b,c,d) if we lived in a world that we knew to be strongly influenced by shock waves from which new information emanated; only for Fig. 7.5(a) can the free boundary evolution be predicted from our models of gas dynamics, and, in this case, the free boundary is stable in the sense that it can evolve temporally in synchronisation with any harmonic wave that impinges on it from either side.

The fact that shock waves can exist as stable coherent free boundaries in gas dynamics has been famous for many years, which is why we were able to introduce them as early as §1.7 of this book. As with most free boundary problems, the computer is now the principal tool for predicting shock wave behaviour, except in certain simple configurations.

Although we have not considered the stability or well-posedness of all the models in §7.1, the examples above illustrate the kind of behaviour that may be encountered. We can summarise by stating that, on the basis of the specialisation to Hele-Shaw flow, Stefan problems appear to be well-posed if neither supercooling nor superheating occurs. Also, a stability analysis of (7.14) and (7.16), as in Exercise 7.20, shows that porous medium free boundaries are well posed when the saturated region is below the free boundary, and ill-posed otherwise, unless the

[144] We have in fact encountered this situation already in Chapter 2 in our discussion of causality.

free boundary is moving in the vertical direction, and even then a saturated region above a free boundary may be stable if the boundary moves down with sufficient speed. However, no general remarks will be made here about flames or contact problems, for which the stability analysis is much more complicated.

The analyses above can all be refined greatly using various asymptotic methods, in particular the methods of weakly nonlinear stability analysis. But none of these refinements can overshadow the ease with which free boundary problems can apparently be ill-posed. We say 'apparently' because our examples may have given the impression that linear instability is synonymous with ill-posedness, but it must be remembered that the relationship between these concepts depends very much on the form of the dispersion relation between the temporal growth rate λ and the wavelength $2\pi/k$. Even if λ is positive for all k, it may depend on k so weakly that the transform procedure mentioned at the end of §7.2.1 may converge for all $t > 0$; this would be the case, for example, for surface gravity waves with $F < 0$.

It is clearly time to make some remarks about the few rigorously justifiable theories that are available for certain classes of free boundary problems. We begin with techniques that may be used to analyse *classical* free boundary problems, i.e. ones in which the field equations and free boundary conditions are to be satisfied pointwise.

7.3 Classical solutions

7.3.1 Comparison methods

Some information about the position of free boundaries and the size of solutions can occasionally be gleaned from comparison methods. The 'stable' one-phase Stefan problem is a case in point. For simplicity, we take $u_m = 0$ and all the other coefficients in (7.1) and (7.2) to be unity. Intuitively, we argue as follows: consider two solutions u_1 and u_2 with $u_1 \geqslant u_2 \geqslant 0$ so that, at some time, $u_{1,2} > 0$ in $\Omega_{1,2}$, respectively, with $\Omega_2 \subset \Omega_1$, and $u_1 > u_2$ in Ω_2 (see Fig. 7.6).

Now suppose that at some later time the expanding free boundary Γ_2 catches up with Γ_1, so that $u_1 = u_2 = 0$ at some point where these free boundaries touch, and where we shall assume they are smooth. From the strong maximum principle for parabolic equations, mentioned in §6.3, the normal velocities of Ω_i, denoted by v_{ni}, satisfy

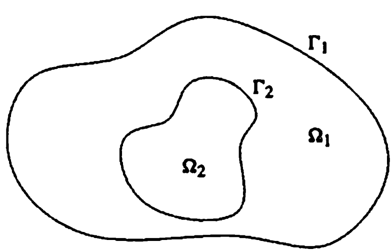


Fig. 7.6 Comparison of free boundaries.

$$v_{n1} = -\frac{\partial u_1}{\partial n} > v_{n2} = -\frac{\partial u_2}{\partial n},$$

contradicting the catching-up assumption. It can be noted that, with the 'unstable' version of the problem when $\Omega_1 \subset \Omega_2$ but $u_1 \leqslant u_2 \leqslant 0$, the argument fails completely, and in these cases the boundaries can cross. This irregular behaviour is consistent with the instability result for the Hele-Shaw problem with suction, which is simply a special case of the supercooled Stefan problem.

7.3.2 Energy methods and conserved quantities

Despite the nonlinearity of free boundary problems, it may be possible to extract some information by more or less straightforward integration. For example, for the porous medium equation in one dimension,

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(u^n \frac{\partial u}{\partial x}\right),$$

with

$$u(x,0) = u_0(x) \quad \text{for } -\infty < x < \infty,$$

it is trivial to show conservation of mass in the form

$$\int_{-\infty}^{\infty} u(x,t)\,\mathrm{d}x = \int_{-\infty}^{\infty} u_0(x)\,\mathrm{d}x, \tag{7.60}$$

and that the centroid is fixed, so that

$$\int_{-\infty}^{\infty} xu(x,t)\,\mathrm{d}x = \text{constant}, \tag{7.61}$$

assuming $u_0(x)$ decays sufficiently rapidly as $|x| \to \infty$. Again, for a Hele-Shaw flow in a region Ω driven by a single source or sink at the origin, an obvious integration of $\nabla^2 p = 0$, with $p = 0$ and $\partial p/\partial n = -v_n$ on the free boundary, and

$$p \sim -\frac{Q}{2\pi}\log r \quad \text{as } r^2 = x^2 + y^2 \to 0, \tag{7.62}$$

gives that the area Ω of the fluid changes at a rate equal to the source strength Q.[145] However, this problem also has an infinite number of conserved quantities

$$\iint_{\Omega} z^m\,\mathrm{d}x\,\mathrm{d}y, \quad m = 1, 2, \ldots, \tag{7.63}$$

where $z = x + \mathrm{i}y$, a fact to which we shall return later.

[145] Note that p is proportional to the Green's function for the fluid domain with Dirichlet boundary conditions.

Sometimes these methods can point to impending disaster. Consider the one-phase one-dimensional Stefan problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } x < s(t), \tag{7.64}$$

with

$$u = 0, \quad \frac{\partial u}{\partial x} = -\frac{ds}{dt} \quad \text{at } x = s(t), \tag{7.65}$$

with supercooled initial data $u(x,0) = u_0(x) < 0$, $0 < x < s_0$, and assume that $\partial u/\partial x\,(0,t) = 0$ for $t > 0$. Integration of the field equation, which is simply an expression of global conservation of heat, gives

$$s(t) + \int_0^{s(t)} u(x,t)\,dx = s_0 + \int_0^{s_0} u_0(x)\,dx. \tag{7.66}$$

Now, if the solution exists for all $t > 0$, then there exists an s_∞ such that $s(t) \to s_\infty \geqslant 0$ and $u(x,t) \to 0$ as $t \to \infty$. Hence, if u_0 is so large and negative that $s_0 + \int_0^{s_0} u_0\,dx < 0$, then we have a contradiction and so we have 'finite-time blow-up', as in §6.6.5. The physical interpretation of this blow-up is that if, in supercooled solidification, the latent heat released at the free boundary is conducted away too quickly because u is too large and negative nearby, then the steep temperature gradient at the front makes it move so fast (by the condition $\partial u/\partial x = -ds/dt$) that the gradient steepens even further and the free boundary moves even faster. If the sign of u is reversed, so that the 'active' phase is liquid above the melting temperature, the free boundary moves in the other direction and instability becomes stability. Moments such as (7.63) can be used to reveal more information about blow-up when $s_0 + \int_0^{s_0} u_0\,dx \geqslant 0$ (see Exercise 7.9).[146]

7.3.3 Green's functions and integral equations

Although Green's functions cannot be used directly to solve free boundary problems, we can sometimes collapse information from the field equation onto the free boundary and hence reduce the problem to a nonlinear integral equation, as in the boundary integral method mentioned at the end of §5.5.2. As an example, consider again the one-phase Stefan problem (7.64) and (7.65) for $-\infty < x < s(t)$, with $u(x,0) = u_0(x)$, assumed to behave suitably as $x \to -\infty$, and where $u_0 \equiv 0$ in the passive phase $x > s(0)$. We define $G(x, \tau - t; \xi)$ such that

$$-\left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial G}{\partial t}\right) = \delta(x-\xi)\delta(t-\tau) \quad \text{for } t < \tau,$$

as usual. However, we make no specification about the boundary behaviour of G, so that the Green's function for the whole line, written as

$$E(x-\xi, \tau - t) = \frac{1}{2\sqrt{\pi(\tau - t)}} e^{-(x-\xi)^2/4(\tau - t)},$$

[146] In fact, the blow-up can occur even if the inequality is not satisfied, as long as u_0 is large and negative near $x = s_0$.

will suffice, as shown in [38]. Then the usual integration in the form

$$\int_0^\tau \int_0^{s(\tau)} \left(u \left(\frac{\partial^2 E}{\partial x^2} + \frac{\partial E}{\partial t} \right) - E \left(\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} \right) \right) \mathrm{d}x \, \mathrm{d}t$$

shows that

$$u(\xi, \tau) = u_0(\xi) * E(\xi, \tau) - \int_0^\tau \frac{\mathrm{d}s(t)}{\mathrm{d}t} E(\xi - s(t), \tau - t) \, \mathrm{d}t, \tag{7.67}$$

where $*$ denotes convolution as usual. When the latent heat is zero the second term is absent and this is simply (6.32). Now, as long as we are careful, we can differentiate with respect to x and take the limit as $\xi \uparrow s(t)$, as in Exercise 7.10, to obtain

$$\frac{1}{2}\frac{\mathrm{d}s}{\mathrm{d}\tau} = -\left.\frac{\partial}{\partial \xi}(u_0 * E)\right|_{\xi = s(\tau)} - \frac{1}{2}\int_0^\tau \frac{\mathrm{d}s(t)}{\mathrm{d}t} \left(\frac{s(\tau) - s(t)}{\tau - t} \right) E\left(s(\tau) - s(t), \tau - t\right) \mathrm{d}t; \tag{7.68}$$

taking the limit directly in (7.67) would yield a less tractable first-kind integral equation. An iterative analysis may now be used to demonstrate existence and uniqueness of the classical solution under appropriate conditions on $u_0(x)$ [38].

Unfortunately, this technique does not readily generalise to multidimensional free boundary problems, a matter to which we will return in §7.5. To make much more progress with such problems, we must be less ambitious and consider generalised rather than classical solutions, but, before we do this, we make one seemingly trivial remark that is often overlooked. We can clearly transform any free boundary problem to one in a fixed domain by a simple change of variable, such as $\xi = x/s(t)$ or $\xi = x - s(t)$ for the one-dimensional Stefan problem. The price we pay is that the field equation then involves coefficients that are global functions of u and hence difficult to analyse. However, such transformations can be of great computational advantage.

*7.4 Weak and variational methods

Since classical free boundary problems are so difficult to analyse rigorously, it is natural to try to ease the mathematical task by demanding a less stringent definition of the concept of solution. One possibility is to follow the ideas introduced in §1.7 and to try to define a weak solution by multiplying the field equation by test functions and integrating in such a way that the free boundary and the conditions imposed on it are automatically incorporated into the integral formulation. Alternatively, we could try to generalise the variational approach described in §5.4 in such a way that the free boundary conditions are automatically satisfied by the minimisers.

In either case the philosophy is the same: we are seeking a formulation which makes sense even in the presence of whatever discontinuities are inherent in the free boundary conditions. The skill comes in writing down a formulation which has good existence and uniqueness properties. Even if this formulation is too unwieldy for us to have much hope of finding explicit formulæ for generalised solutions, we may hope that numerical discretisations can be devised that could be proved to tend to the weak or variational solution when the appropriate step size decreases.

In Chapter 1 we have already seen the dangers of this philosophy. As we reminded ourselves in §1.7, when we used (1.22) to define weak solutions of hyperbolic equations, we found we had cast our net too widely and we needed extra information before we could hope for uniqueness. In this chapter we are confronted with a far wider range of problems, in which the field equation may be elliptic, parabolic or hyperbolic, and for which all manner of free boundary conditions may be prescribed. Hence, we have a major theoretical task ahead of us, and we begin by making two general observations.

First, it will come as no surprise if we announce that the chances of finding a generalised formulation of an arbitrary free boundary problem are very small,[147] even though a surprising number of practically relevant problems are suitable cases for treatment. Second, a crucial diagnostic feature is the 'strength' of the conditions at the free boundary, i.e. how many derivatives, if any, the solution may be expected to have there. This especially concerns variational approaches because we have a chance of formulating the problem as in §5.4 when, and only when, the free boundary conditions are benign enough that only the highest derivatives in the field equation are discontinuous. We saw there that the Euler–Lagrange equation is of one order higher than that of the derivatives appearing in the Lagrangian; hence a minimisation problem might still make sense and have the above-mentioned degree of smoothness. We begin by considering this desirable situation.

7.4.1 Variational methods

The simplest situation of all arises when the free boundary emerges directly from the kind of variational statement of the type used in §5.4. Suppose, for example, that we wish to solve the vorticity problem (7.39). Replacing ψ by u and letting $u > 0$ in the patch of vorticity, it is natural to consider

$$\min \int \left(\frac{1}{2} |\nabla u|^2 + F(u) \right) \mathrm{d}\mathbf{x}, \tag{7.69}$$

where

$$\frac{\mathrm{d}F}{\mathrm{d}u} = \begin{cases} -f(u), & u > 0, \\ 0, & u < 0, \end{cases}$$

and the minimisation is taken over suitable test functions satisfying the required conditions at fixed boundaries. To prove rigorously that this variational statement has a unique solution is still hard work, even when, say, f is differentiable and $f(0) = 0$. However, for our purposes, it is even more important that the variational statement achieves the objectives of any generalised solution, namely that it includes all classical solutions. Hence we need to show that, if a free boundary exists at which $u = 0$, then the minimiser has the correct behaviour there, with $\partial u/\partial n$ being continuous. We can see this by assuming that we can write the free

[147] We recall that we could only define weak solutions of hyperbolic equations when they were in conservation form.

boundary as $S(\mathbf{x}) = 0$, with $S > 0$, where $u > 0$, and splitting (7.69) into contributions from either side. When we change (u, S) to $(u + \delta u, S + \delta S)$ we obtain, to lowest order,

$$\int_{u>0} \delta u(\nabla^2 u + f(u))\,\mathrm{d}\mathbf{x} + \int_{u<0} \delta u \nabla^2 u\,\mathrm{d}\mathbf{x}$$
$$+ \int_{S=0} \left(\left[\frac{\partial u}{\partial n} \right]_{S=0-}^{S=0+} \delta u - \left[\frac{1}{2} |\nabla u|^2 + F(u) \right]_{S=0-}^{S=0+} \frac{\delta S}{|\nabla S|} \right) \mathrm{d}s. \quad (7.70)$$

taking the normal direction into the region where $u > 0$. The first two terms vanish by (7.39), and F is continuous; hence, since δu and δS are independent, u and $\partial u / \partial n$ are continuous.

An even more subtle situation arises when we try the same idea on the obstacle problem (7.40)–(7.42). Since the obstacle topography f does not in general satisfy $\nabla^2 f = 0$, we cannot expect to get a solution by minimising the Dirichlet integral $\frac{1}{2} \int |\nabla u|^2\,\mathrm{d}\mathbf{x}$ over all small, smooth-enough perturbations. But it is natural, physically and mathematically, to consider the *restricted* or *unilateral* minimisation problem

$$\min_{u \geqslant f} \int \frac{1}{2} |\nabla u|^2\,\mathrm{d}\mathbf{x}, \quad (7.71)$$

together with suitable conditions on u at the boundary of the membrane.

Again, it transpires that this minimisation can indeed yield an acceptable generalised solution to the obstacle free boundary problem just as long as u and ∇u are continuous at the free boundary, but now the proof requires quite a different idea. In particular, it relies on the interesting identification between the following three concepts.

1. The unilateral minimisation statement (7.71).
2. The so-called *linear complementarity* statement

$$\nabla^2 u \leqslant 0, \quad u \geqslant f, \quad (\nabla^2 u)(u - f) = 0. \quad (7.72)$$

3. The so-called *variational inequality* over the fixed domain Ω containing both the 'contact' and 'non-contact' regions:

$$\int_\Omega \nabla u \cdot \nabla(v - u)\,\mathrm{d}\mathbf{x} \geqslant 0 \quad \text{for all } v \geqslant f, \quad (7.73)$$

assuming that v satisfies suitable fixed boundary conditions.

It is difficult to establish the equivalence between the original problem (7.40)–(7.42) and (7.71)–(7.73) rigorously, although (7.40)–(7.42) clearly implies the linear complementarity statement (7.72) (note that, when the membrane is in contact with the obstacle, its linearised curvature $\nabla^2 u$ must be negative). To relate the unilateral minimisation (7.71) and the variational inequality (7.73), note that, if

u_1 and u_2 are candidates for (7.71), then so is $\epsilon u_1 + (1-\epsilon)u_2$ for any ϵ such that[148] $0 \leqslant \epsilon \leqslant 1$. Hence the minimiser is such that

$$\int_\Omega |\nabla u|^2 \, \mathrm{dx} \leqslant \int_\Omega |\nabla (\epsilon u_1 + (1-\epsilon)u_2)|^2 \, \mathrm{dx}.$$

For small ϵ, the right-hand side is equal to

$$\int_\Omega |\nabla u_2|^2 \, \mathrm{dx} + 2\epsilon \int_\Omega \nabla u_2 \cdot \nabla (u_1 - u_2) \, \mathrm{dx} + O(\epsilon^2),$$

and so, with $u_2 = u$ and $u_1 = v$,

$$\int_\Omega \nabla u \cdot \nabla (v - u) \, \mathrm{dx} \geqslant 0,$$

and the converse is obvious since $|\nabla u|^2 - |\nabla v|^2 = -2\nabla u \cdot \nabla(v-u) - |\nabla(u-v)|^2$. Further details can be found in [17], where the advantages of each of the reformulations (7.71)–(7.73) over the original problem (7.40)–(7.42) are also explained in detail; in particular, the variational inequality is a good starting point for existence and uniqueness results, and the unilateral minimisation and the linear complementarity formulation can have easy numerical implementations, as we will see shortly.

This kind of approach only works if, firstly, the field equations are Euler–Lagrange equations of a well-behaved functional and, secondly, the free boundary conditions are sufficiently smooth. We have mentioned the latter fact several times already and it is clear that this smoothness is needed to relate, say, (7.72) to (7.73). Hence, if we are faced with a free boundary problem lacking the requisite smoothness, it is tempting to try to transform it to one with better 'regularity'.

A famous class of problems susceptible to such smoothing are the flows of a liquid in a porous medium modelled by the steady versions of (7.14)–(7.16). For these problems we can simply integrate with respect to the vertical independent variable y (the *Baiocchi transformation*) and write the free boundary as $y = h(x)$ to give

$$\int_y^h p(x, y') \, \mathrm{d}y' = u(x, y).$$

We restrict ourselves to stable situations where the saturated part of the porous medium lies below the dry part. Then $u \geqslant 0$ in the saturated region because $p \geqslant 0$, and, extending u and p to be zero in the dry region, a simple calculation gives

$$\nabla^2 u = \begin{cases} 1, & y < h, \\ 0, & y > h, \end{cases} \qquad \text{with} \qquad u = \frac{\partial u}{\partial n} = 0 \quad \text{at } y = h. \tag{7.74}$$

We thus have a free boundary problem which can be reformulated as a linear complementarity problem similar to (7.72), or as the variational inequality

[148] Functional analysts say that the minimisation is over a convex cone in a suitable function space.

$$\int_\Omega \nabla u \cdot \nabla(v-u)\,\mathrm{d}\mathbf{x} \geqslant \int_\Omega (u-v)\,\mathrm{d}\mathbf{x}, \tag{7.75}$$

where the test functions $v(x,y)$ are positive and satisfy appropriate fixed boundary conditions, and Ω is the entire porous medium, including both saturated and unsaturated regions.

An even more striking example of a smoothing transformation concerns the relationship between the one-phase Stefan problem (7.64) and (7.65) and the oxygen consumption problem (7.27). For the latter, we have seen that a physical argument of mass balance at the interface suggests the free boundary conditions

$$c = \frac{\partial c}{\partial x} = 0 \quad \text{at } x = s(t). \tag{7.76}$$

Now suppose oxygen is being consumed in $0 < x < s(t)$, with $\partial c/\partial x = 0$ at $x = 0$ and $s(0) = 1$. The condition (7.76) gives enough smoothness for the complementarity formulation

$$c\left(\frac{\partial c}{\partial t} - \frac{\partial^2 c}{\partial x^2} + 1\right) = 0, \quad c \geqslant 0, \quad \frac{\partial c}{\partial t} - \frac{\partial^2 c}{\partial x^2} \geqslant -1 \tag{7.77}$$

to be shown to be equivalent to what is called an *evolution variational inequality*

$$\int_0^1 \frac{\partial c}{\partial t}(v-c)\,\mathrm{d}t + \int_0^1 \frac{\partial c}{\partial x}\frac{\partial}{\partial x}(v-c)\,\mathrm{d}x \geqslant \int_0^1 (-1)(v-c)\,\mathrm{d}x \tag{7.78}$$

for all positive $v(x,t)$ satisfying appropriate conditions at $x = 0, 1$; with our particular initial and boundary data, we just need $v = 0$ at $x = 1$. Again, it is difficult to prove the equivalence of (7.27), (7.77) and (7.78), although (7.77) and (7.78) can be used to prove existence and uniqueness of weak solutions. But what is interesting is that the Stefan problem (7.64) and (7.65) does not have the requisite smoothness for a complementarity or variational formulation, yet it is apparently intimately related to (7.27): when we formally set

$$\frac{\partial c}{\partial t} = u \tag{7.79}$$

in (7.27), we obtain

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}. \tag{7.80}$$

Moreover, since

$$c = \frac{\partial c}{\partial x} = \frac{\partial c}{\partial t} = 0 \quad \text{on } x = s(t),$$

we see that $u = 0$ on $x = s(t)$ and, since

$$\frac{\partial^2 c}{\partial x^2}\frac{\mathrm{d}s}{\mathrm{d}t} + \frac{\partial^2 c}{\partial x\,\partial t} = 0 \quad \text{on } x = s(t),$$

we find

$$\frac{\partial u}{\partial x} = -\frac{\mathrm{d}s}{\mathrm{d}t} \quad \text{on } x = s(t),$$

which is (7.64) and (7.65). Hence the Stefan problem is, formally, the time derivative of the oxygen consumption problem, and this applies in higher dimensions as well. However, there is a catch here because we can only use (7.80) on the *unconstrained* version of (7.27) in which the sign of c is not specified, rather than the nicely-behaved constrained problem that can be written as (7.77) and (7.78). In fact, we have indicated that, in two space dimensions, the Stefan problem can behave very badly, and hence so can the unconstrained oxygen consumption problem. Yet the constrained oxygen consumption problem has perfectly good existence, uniqueness and continuous-dependence properties, at least as far as weak solutions are concerned.

The complementarity approach can also be applied to some other models described in the introduction, such as the dead core model mentioned in §5.11.3. Equally, for the American option model a simple calculation shows that

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \leqslant r\left(V - S\frac{\partial V}{\partial S}\right). \tag{7.81}$$

In financial terms, this is a natural requirement which says that the return on a perfectly hedged option portfolio (the left-hand side) is never more than the risk-free rate on the same portfolio (the right-hand side). If it is optimal to hold the option, we have equality, but if we are in the exercise region it is suboptimal to hold the option and so the return on the hedged portfolio is less than the risk-free rate. This inequality can be used with the condition $V \geqslant \max(E - S, 0)$ and the fact that, whatever S, one or other of these inequalities is an equality, to make a linear complementarity statement [47]. This in turn can be shown to imply the smoothness condition that $\partial V/\partial S$ is continuous at the free boundary, i.e. the classical free boundary conditions (7.20) hold (see Exercise 7.5).[149]

As our final example, we return to the smooth elastic contact problems mentioned in §7.1.3.2. Amazingly, they can all be formulated as variational inequalities because, first, the elastic energy is minimised over the three-dimensional bulk of the contacting solids, the Euler–Lagrange equations being (4.73) with $\partial/\partial t = 0$; second, the normal displacement and tractions are continuous (the tangential traction is zero) on the contact region, which to lowest order lies in a prescribed surface S and is bounded by the one-dimensional free boundary Γ; third, no traction is applied on S outside the constant region. The reason that this is a variational inequality is now clear: on one side of Γ, conventional boundary conditions are prescribed and, on the other side, the no-traction condition is the natural boundary condition [12] for the elastic energy. Put crudely, the variational inequality is clever enough to select that curve Γ on which the boundary conditions can switch from contact conditions to natural boundary conditions with the requisite smoothness, just as it does for the American option.

In summary, it is a red-letter day when a free boundary problem can be cast into a variational formulation. Not only does this usually mean that rigorous statements can be made about existence and uniqueness of a generalised solution, but the numerical calculation of this solution can be relatively trivial. For example, for the vector linear complementarity problem

[149]In this respect the no-arbitrage condition acts rather like the second law of thermodynamics.

$$\mathbf{u} \geqslant \mathbf{0}, \quad \mathbf{A}\mathbf{u} \geqslant \mathbf{b}, \quad \mathbf{u}^{\mathsf{T}}(\mathbf{A}\mathbf{u} - \mathbf{b}) = 0, \tag{7.82}$$

where positivity of a vector $\mathbf{u}$ means positivity of each component u_i, we can put

$$\bar{u}_i^{k+1} = \frac{1}{a_{ii}}\left(b_i - \sum_{j<i} a_{ij}u_j^{k+1} - \sum_{j>i} a_{ij}u_j^k\right), \tag{7.83}$$

where $\mathbf{A} = (a_{ij})$, and introduce a relaxation parameter ω so that

$$u_i^{k+1} = \max\left(0, u_i^k + \omega(\bar{u}_i^{k+1} - u_i^k)\right)$$

gives an easily implemented updating for u_i^{k+1} in terms of u_i^k, as described in more detail in [13, 17, 47]. Hence, for example, option valuation can be carried out quickly enough for the traders on Wall Street.

Despite all the success stories mentioned above, there remain many more free boundary problems where there is inevitably insufficient smoothness for us to be able to proceed variationally. Sometimes, however, we can have recourse to the following method.

7.4.2 The enthalpy method

The enthalpy approach to the prototypical Stefan problem is similar in spirit to the idea of weak solutions to hyperbolic equations. It is based on thinking of the free boundary condition (7.4) as a Rankine–Hugoniot relation for a partial differential equation in conservation form. Indeed, we recall that the one-dimensional heat equation

$$\rho c\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}$$

was derived from the integrated form of the conservation law

$$\frac{\partial}{\partial t}(h(u)) + \frac{\partial}{\partial x}\left(-k\frac{\partial u}{\partial x}\right) = 0, \tag{7.84}$$

where $h = \rho c u$ is the heat content or *enthalpy* of the material. We now observe that, if we set the melting temperature u_m to be zero, without loss of generality, and write

$$h(u) = \begin{cases} \rho c u, & u < 0 \quad \text{(solid)}, \\ \rho c u + \rho L, & u > 0 \quad \text{(liquid)}, \end{cases} \tag{7.85}$$

then the Rankine–Hugoniot condition for (7.84) is $[-k\,\partial u/\partial x]_{\text{solid}}^{\text{liquid}} = \rho L\,\mathrm{d}s/\mathrm{d}t$, in accordance with (7.4). However, the other free boundary condition $u = 0$ is not a Rankine–Hugoniot condition but rather the label of the temperature at which the nonlinearity occurs; it indicates that the Stefan problem is really a limiting nonlinear diffusion problem.

To make a mathematical theory out of these ideas, we begin by dealing just with continuous functions, replacing (7.85) by $u = F(h)$, where

$$F(h) = \begin{cases} h/\rho c, & h < 0, \\ 0, & 0 < h < \rho L, \\ h/\rho c - L/c, & \rho L < h, \end{cases} \tag{7.86}$$

and regarding (7.84) as an equation for $F(h)$. Then we need to interpret (7.84) properly, given that $\partial u/\partial x$ and $\partial^2 u/\partial x^2$ do not exist at the free boundary. To do this, we simply multiply (7.84), with u regarded as a function of h, by a test function v which, in the simplest case of the Cauchy problem on $-\infty < x < \infty$, vanishes for sufficiently large x and at $t = T$, and is infinitely differentiable. Hence we can motivate the definition of a weak solution of this Stefan problem as a pair of functions u and h satisfying (7.86) and such that

$$\int_0^T \int_{-\infty}^{\infty} \left(h\frac{\partial v}{\partial t} + ku\frac{\partial^2 v}{\partial x^2} \right) \mathrm{d}x\, \mathrm{d}t = \int_{-\infty}^{\infty} h\big|_{t=0} v(x,0)\, \mathrm{d}x \tag{7.87}$$

for all such test functions. The argument leading to (1.27) can then be used to prove that any such u satisfies (7.3) and (7.4) (with $u_m = 0$) at a phase boundary, where $\partial u/\partial x$ is discontinuous. Moreover, this pair of functions (u, h) can be proved to exist and to be unique (see [17]). The general idea of the proof is a by-product of the most important practical implication of (7.87), which is the result that the temporal discretisation

$$\frac{h_{n+1} - h_n}{\delta t} = \frac{\partial}{\partial x}\left(k\frac{\partial u_n}{\partial x} \right), \quad u_{n+1} = F\left(h_{n+1}\right), \tag{7.88}$$

which is easily implemented, converges to the unique weak solution. It is interesting to note that the convergence of the analogous discretisation for hyperbolic equations cannot be proved; this is because the weak solution is then not unique. In fact, it is, in general, much more difficult to implement discretisations of the weak formulation of hyperbolic conservation laws than it is for diffusive ones. We recall that in the former case all the free boundary conditions are Rankine–Hugoniot conditions for conservation laws, whereas in the Stefan problem the condition $[k\nabla u \cdot \mathbf{n}]_{\text{solid}}^{\text{liquid}} = -\rho L v_n$ is triggered by the discontinuity in the enthalpy. The rapid transition across, say, a gasdynamic shock wave can easily generate spurious high-frequency oscillations in the solution of the discretised weak formulation, and these oscillations can propagate as waves throughout the flow more easily than can the oscillations generated at a phase boundary by the solution of (7.88).

The algorithm (7.88) is known as the *enthalpy method* for the Stefan problem. It is a simple matter to generalise the method to more than one space dimension, even when it comes to proving existence and uniqueness of the weak solution; this is in contrast to the classical solution whose existence is hard to prove in more than one dimension.

We see the most dramatic implication of the enthalpy method when we enquire about the relationship between classical and weak solutions of the Stefan problems.

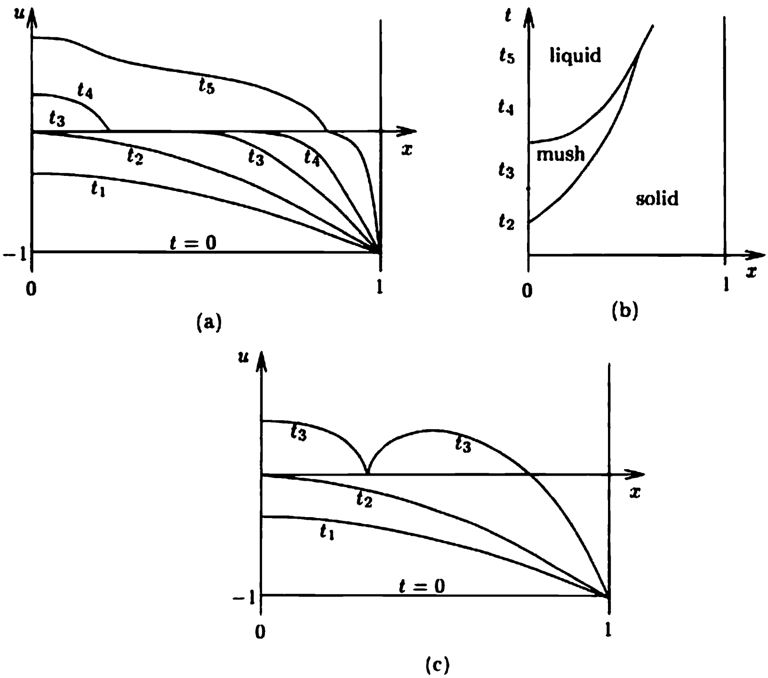


Fig. 7.7 The Stefan problem with volumetric heating.

Although, as stated above, we can prove that, when any two regions in which u is infinitely differentiable are joined by a free boundary where (7.3) and (7.4) hold, we have a weak solution, it is not the case that such free boundaries are the only ones that can appear in weak solutions. Indeed, it can be proved that, if we consider the slightly modified problem with a volumetric heat source[150]

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + Q \quad \text{for } 0 < x < 1, \tag{7.89}$$

with

$$u(x,0) = -1, \quad \frac{\partial u}{\partial x}(0,t) = 0, \quad u(1,t) = -1, \tag{7.90}$$

then, by the enthalpy method, for $Q > 2$ the solution evolves as in Fig. 7.7(a).

This observation is most easily made by plotting the output for the appropriate modification of the algorithm (7.88). We see the appearance of a so-called *mushy region* in which $u \equiv 0$ and h increases for $t_1 < t < t_2$. The Stefan free boundary conditions (7.3) and (7.4) are satisfied on neither of the boundaries of this mushy

[150]This is a simple model for melting by constant volumetric heating, say in resistance welding: two metal sheets in $-1 < x < 0$ and $0 < x < 1$ are to be welded along the x axis with their outer faces $x = \pm 1$ held at the initial temperature -1.

region; only after the mushy region has vanished (Fig. 7.7(b)) does such a classical Stefan free boundary emerge, for example at $t = t_5$.

Additionally, it can be shown that, if we were to insist that the only free boundary in the problem was to be one at which the classical Stefan conditions (7.3) and (7.4) were satisfied, then the solution would evolve quite differently, and, as in Fig. 7.7(c), it would involve a superheated region in which the solid temperature exceeds the melting temperature. We recall that our stability argument for the case of zero specific heat in §7.2.3 suggested that such superheating yields a free boundary that is unstable to small disturbances in two dimensions.

The physical interpretation of Fig. 7.7(a), with its prediction of coexisting liquid and solid phases, demands much more detailed physical scrutiny than we can give here, but it is likely that Fig. 7.7(a) is more realistic than Fig. 7.7(c) in many practical situations. What has happened mathematically is that the enthalpy formulation has legislated against superheating (as it would against supercooling) because of the nature of the function $h(u)$ in (7.85). We remember that at no point in our classical formulation did we prohibit superheating or supercooling and the weak formulation is a valuable reminder that these constraints are just as important as they were in the oxygen consumption problem.

Although the Stefan model is one of the commonest free boundary problems where a weak formulation is available, the idea can be tried on any problem in divergence form. For example, when applied to the one-dimensional porous medium equation (6.72), it gives

$$u = 0, \quad u^{n-1}\frac{\partial u}{\partial x} = -\frac{\mathrm{d}s}{\mathrm{d}t} \tag{7.91}$$

as free boundary conditions for a weak solution, the second of which can sometimes be interpreted as conservation of mass at the free boundary. However, we must remind the reader of the fragility of the situation. It only needs, say, the melting temperature or latent heat to be a function of position for the weak formulation of the Stefan problem to cease to be available.

Free boundary problems are one area of partial differential equations theory where rigorous existence and uniqueness proofs and justifiable numerical algorithms far outnumber the techniques available for finding explicit solutions. Some of the few techniques that are available are described below.

7.5 Explicit solutions

Only two of the techniques described in Chapters 4–6 can be used with confidence to find explicit solutions of free boundary problems. They are, firstly, the use of similarity variables which, we recall, includes travelling waves, and, secondly, the use of complex variables if the field equation happens to be Laplace's equation or, perhaps, the biharmonic equation. The use of complex variables leads to the idea of conformal invariance and thus can be regarded as a special case of a similarity method, but the associated techniques are distinctive enough to merit separate discussion.

7.5.1 Similarity solutions

All the remarks about similarity and group invariance made in §6.5 apply to any differential equation, with a free boundary or otherwise. However, success always relies on identifying the relevant group, so we will simply list some informative examples.

7.5.1.1 Travelling waves: similarity variable $x - Vt$

When the independent variables x and t do not appear explicitly in the field equations we can always seek solutions that depend just on $x - Vt$ for some constant V and, for certain one-dimensional free boundary problems, this allows the free boundary to be at $x = Vt$. For example, the one-phase Stefan problem (7.64) and (7.65), in which water occupies the region $x < Vt$, has a solution $u = F(x-Vt)$ as long as

$$\frac{\mathrm{d}^2F}{\mathrm{d}\xi^2} + V\frac{\mathrm{d}F}{\mathrm{d}\xi} = 0, \tag{7.92}$$

where, since the ice in $x > Vt$ has temperature $u = 0$,

$$F(0) = 0, \quad \frac{\mathrm{d}F}{\mathrm{d}\xi}(0) = -V. \tag{7.93}$$

Hence

$$u(x,t) = -1 + \mathrm{e}^{-V(x-Vt)}, \tag{7.94}$$

and we see that, if $V > 0$, and we have what we regarded in §7.3.2 as a stable situation with the water temperature above zero, then $u \to +\infty$ as $x - Vt \to -\infty$. However, if $V < 0$, then $u \to -1$ as $x - Vt \to -\infty$ and we will see the significance of this limiting value shortly.

Travelling wave solutions can be sought for many of the other models described in §7.1. At one extreme, the Rankine–Hugoniot equations themselves provide very simple travelling wave solutions for shocks in hyperbolic conservation laws, while the problem for travelling two-dimensional surface gravity waves is still a challenging task for numerical and mathematical analysts, unless the wave slope is small.

In combustion theory, our indeterminacy after (7.26) can be resolved by modelling flames travelling into a premixed environment with (7.21), (7.22) and (7.26), and, after making the key assumption that we can neglect the reaction ahead of the flame, seeking a travelling wave. We find the profiles

$$T(x,t) = T_A + (T_0 - T_A)\mathrm{e}^{-V(x-Vt)/k}, \quad c(x,t) = c_0\left(1 - \mathrm{e}^{-V(x-Vt)/D}\right) \tag{7.95}$$

for the temperature and concentration, respectively, in $x > s(t) = Vt$, where T_A and c_0 are, respectively, the values of the temperature and concentration far ahead of the flame. The compatibility condition (7.25), when applied to (7.95) for small positive values of $x - Vt$, shows that T_0 is given by $T_0 - T_A = Ac_0/\alpha$, which has the simple physical interpretation that the rise in temperature is that achieved by the complete reaction. Finally, (7.26) gives the flame velocity as $V = \sqrt{2\alpha/DE}\left(kT_0/(T_0 - T_A)\right)\mathrm{e}^{-E/2T_0}$.

Unfortunately, this approximate solution loses self-consistency at large distances ahead of the flame because the reaction terms in (7.23), although small, dominate the derivatives unless T_A is absolute zero. This is the so-called 'cold boundary difficulty'. The contradiction can be explained away by the fact that, in real life, the model only applies for finite times and in bounded regions. Indeed, if we apply the condition $T = T_A$ at a finite, but suitably large, distance ahead of the flame, (7.95) gives a good approximation to T and $\partial T/\partial x$ except, possibly, near this cold boundary.

There is a famous travelling wave solution of the equations (2.3) and (2.4) of unsteady one-dimensional gas dynamics, where we recall that the free boundary conditions at a shock moving with speed V are given by (2.49). We examine the flow produced by instantaneously pushing a piston with speed V_p into a tube containing gas at rest at pressure p_0 and density ρ_0. By seeking a solution in which u, p and ρ are all constants for $V_p t < x < Vt$, it is relatively easy to see that the free boundary velocity is

$$V = \frac{\gamma+1}{4}\left(V_p + \sqrt{V_p^2 + \frac{16a_0^2}{(\gamma+1)^2}}\right),$$

where $a_0^2 = \gamma p_0/\rho_0$ (see Exercise 7.12). Similarity solutions also exist in multi-dimensional steady flow past wedges and cones, where non-uniqueness can occur (non-existence of the solution is also possible).

7.5.1.2 Other similarity variables

With a particular class of initial and boundary conditions, the Stefan problem provides an informative reduction of a partial differential equation free boundary problem to one for an ordinary differential equation.[151] It is easy to see that, as long as the initial temperatures on either side of $x = s(t)$ are constant, with one phase in $s(t) < x < \infty$ and the other in $-\infty < x < s(t)$, the initial and free boundary conditions are all invariant under the transformation $x' = e^{\lambda}x$, $t' = e^{2\lambda}t$, $s' = e^{\lambda}s$, $u' = u$. Hence, as in (6.44), we can set

$$s(t) = \alpha t^{1/2}, \quad \eta = xt^{-1/2} \quad \text{and} \quad u = U(\eta), \tag{7.96}$$

where α is a constant, to give

$$\frac{d^2U}{d\eta^2} + \frac{1}{2}\eta\frac{dU}{d\eta} = 0. \tag{7.97}$$

In the one-phase case with water in $-\infty < x < s(t)$, and $u(x,0) = u_0 =$ constant and $s(0) = 0$, we find

$$U(\alpha) = 0, \quad \frac{dU}{d\eta}(\alpha) = -\frac{\alpha}{2},$$

and hence, from (7.97),

[151] We recall that a similar group invariance argument was used to find some free boundaries for the porous medium equation in §6.6.

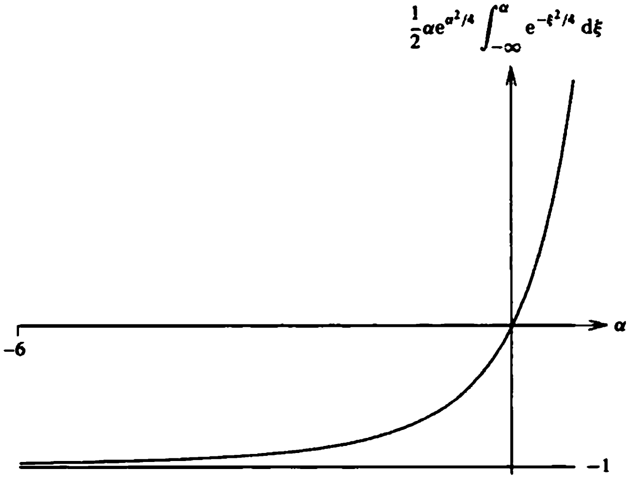


Fig. 7.8 Solving (7.98).

$$\frac{\alpha}{2}e^{\alpha^2/4}\int_{-\infty}^{\alpha} e^{-\eta^2/4}\,\mathrm{d}\eta = u_0. \tag{7.98}$$

This transcendental equation for α can be shown to have a unique solution when $u_0 > -1$ (see Exercise 7.1), and equations like it have found many applications ranging from estimates for the time necessary to freeze food to the valuation of American options. However, from Fig. 7.8, it is easy to see that no real solution exists when $u_0 \leqslant -1$ which, like the argument after (7.66), shows that, even in one space dimension, only a certain amount of supercooling can be tolerated. This critical value of u_0 is that which just allows the existence of the travelling wave (7.94).

*7.5.2 Complex variable methods

Complex variable methods can be used to advantage on many free boundary problems for Laplace's equation. For example, they have been particularly effective in studying the class of gravity-free, inviscid, irrotational, steady flows introduced as Helmholtz flows satisfying (7.36). The key point about these flows is that the free boundary conditions can be written just in terms of the gradient of the potential ϕ. Hence we can write these conditions trivially in terms of the *complex velocity potential* $w(z) = \phi + i\psi$, where $z = x + iy$, as

$$\psi = \text{constant} \quad \text{on} \quad \left|\frac{\mathrm{d}w}{\mathrm{d}z}\right| = 1. \tag{7.99}$$

Thus, if we work in the hodograph plane[152] of the complex variable $\mathrm{d}w/\mathrm{d}z = u - \mathrm{i}v$, then we simply have that w, which is an analytic function of z and hence of

[152]This term is now introduced for reasons quite different from those in Chapter 4; note that the axes are now u and $-v$.

$\mathrm{d}w/\mathrm{d}z$, has constant imaginary part on part of the unit circle. If this Dirichlet problem for $\psi(u,v)$ can be solved, then its solution gives a functional relation between w and $\mathrm{d}w/\mathrm{d}z$, that is, an ordinary differential equation which reduces to a quadrature. Unfortunately, the Dirichlet problem also involves the conditions on any fixed boundaries that may be present in the flow, and these usually become unmanageable when written in terms of $u - \mathrm{i}v$. However, this disaster is avoided when any such fixed boundaries are straight. This means that, when we work with $\log \mathrm{d}w/\mathrm{d}z = W(z)$, say, so that the free boundary is $\Re W = 0$, then $\Im W = -\tan^{-1}(v/u)$ is constant on the fixed boundaries. Thus the region of the W plane corresponding to the fluid flow has a polygonal boundary. But, since ψ is constant on both the free and fixed boundaries, the same statement can be made about the geometry of the flow domain in the w plane, which is often just a half-plane. Hence we can resort to the general Schwarz–Christoffel map to relate W to w; this map is a particular explicit realisation of the powerful Riemann mapping theorem, which states that there is a unique conformal map between simply-connected regions in two complex planes when three conditions are satisfied: the interiors of the regions must map into each other, the boundaries must map into each other, and three pairs of boundary points must be identified with each other (see [16]). The details can be quite intricate (see Exercises 7.14 and 7.15), but the key idea is simple: it is the uniqueness of the mapping that pins down $W(z)$, and hence $\mathrm{d}w/\mathrm{d}z$, as a function of w. In this way, an extraordinary variety of flows involving single and multiple jets and cavities can be constructed, subject only to the restriction that the fixed boundaries are straight; see [4] for a comprehensive description of this technique.

Almost as effective is the application of conformal maps to another class of steady flows, namely those through porous media in the presence of gravity, as modelled by (7.14) and (7.16). Now the free boundary conditions are

$$p = 0, \tag{7.100}$$

$$|\nabla p|^2 + \frac{\partial p}{\partial y} = 0, \tag{7.101}$$

which again suggests consideration of the hodograph plane $\mathrm{d}w/\mathrm{d}z = u - \mathrm{i}v$, where now $(u,v) = -\nabla(p + y)$. The facts that (7.101) states that $u^2 + (v + 1/2)^2 = 1/4$ on the free boundary, and that the flow direction is $-\arg \mathrm{d}w/\mathrm{d}z$, mean that the flow domain in the hodograph plane is now bounded by a circle and straight lines. This demands more ingenuity when mapping onto the w plane but many examples can be solved this way.

It is unfortunate that so few evolution problems are susceptible to conformal mapping methods. A little progress can be made with one irrotational inviscid gravity-free flow (see Exercise 7.16) but things are much better for Hele-Shaw flows. Since these flows are limiting cases of Stefan problems, for which there are few multidimensional solutions, almost anything we can say in two dimensions is of great practical value.

The power of the method is most apparent when we consider the problem of blowing or sucking fluid from a single point, which is the simplest possible driving mechanism in practice, and can be realised by mounting a hypodermic syringe in one of the plates of the cell. What we shall be able to do will prove far more valuable than the Green's function reductions, such as (7.67). That method allowed us to collapse the problem onto the free boundary, but only in the form of a 'global' integro-differential equation that was to be satisfied there, and from which it is difficult to find any useful explicit solutions. Using complex variables, we can write one of the free boundary conditions as a local differential relation for the mapping function. Although the problem is still global in nature because of the relationship between the real and imaginary parts of an analytic function, luckily the Hele-Shaw problem has so much 'structure' that we can find a huge range of explicit solutions.

The argument goes as follows. Assume we can map the boundary of the fluid region Ω univalently (i.e. one-to-one) onto the unit circle in an auxiliary ζ plane, say with $z = f(\zeta, t)$, with the source or sink at $z = 0$ being mapped to $\zeta = 0$. Then the problem in the ζ plane is to find a harmonic function p vanishing on the unit circle and with the singularity

$$p \sim -\frac{Q}{2\pi} \log|\zeta| \quad \text{as } |\zeta| \to 0,$$

where Q is the strength of the source. Hence

$$p = -\Re \frac{Q}{2\pi} \log \zeta, \tag{7.102}$$

and we simply need to find f in order to be able to relate ζ back to z. To do this we recall the kinematic condition (7.11), which requires that

$$\frac{\partial p}{\partial t} - |\nabla p|^2 = 0 \tag{7.103}$$

on the free boundary. Now

$$-\frac{2\pi}{Q}\frac{\partial p}{\partial t} = \Re \frac{1}{\zeta}\frac{\partial \zeta}{\partial t}\bigg|_{z \text{ fixed}} = -\Re \frac{1}{\zeta}\frac{\partial f/\partial t}{\partial f/\partial \zeta}; \tag{7.104}$$

also, on $|\zeta| = 1$,

$$\frac{4\pi^2}{Q^2}|\nabla p|^2 = \frac{1}{|\zeta|^2}\left|\frac{\mathrm{d}\zeta}{\mathrm{d}z}\right|^2 = \frac{1}{|\partial f/\partial \zeta|^2}. \tag{7.105}$$

Thus we arrive at our unconventional differential equation for f in the form

$$\Re\left(\zeta \overline{\frac{\partial f}{\partial \zeta}} \frac{\partial f}{\partial t}\right) = \frac{Q}{2\pi} \tag{7.106}$$

on $|\zeta| = 1$, with f being analytic in $|\zeta| < 1$. This formulation enables many free boundary explicit solutions of the Hele-Shaw problem to be written down. For example, with

$$f = a_1(t)\zeta + a_2(t)\zeta^2 \tag{7.107}$$

and a_i real, without loss of generality, by equating coefficients on $|\zeta| = 1$, we find that

$$a_1 \frac{\mathrm{d}a_1}{\mathrm{d}t} + 2a_2 \frac{\mathrm{d}a_2}{\mathrm{d}t} = \frac{Q}{2\pi},$$
$$a_1 \frac{\mathrm{d}a_2}{\mathrm{d}t} + 2a_2 \frac{\mathrm{d}a_1}{\mathrm{d}t} = 0,$$

and hence

$$a_1^2 + 2a_2^2 = \frac{Qt}{\pi} + \text{constant}, \quad a_1^2 a_2 = \text{constant};$$

a sequence of free boundaries in which a limaçon at $t = 0$ terminates in a cardioid at $t = t_2$ is shown in Fig. 7.9.

Experimentation with such polynomial maps other than multiples of the identity suggests that any sequence of contracting (suction) free boundaries generated by these maps terminates before all the fluid is extracted, forming a cusp in the free boundary, and indeed this can be proved to be true. Perhaps this is not so surprising since we know that 'contracting' Hele-Shaw flows are known to be linearly unstable from §7.2.3. What we now see is that nonlinearity seems to do little to ease their fate and, on the contrary, it engenders finite-time blow-up.

We can now note that the moment conservation (7.63) is a simple deduction from (7.105) since

$$\frac{\mathrm{d}}{\mathrm{d}t} \iint_\Omega z^m \,\mathrm{d}x\,\mathrm{d}y = \int_{\partial\Omega} z^m v_n \,\mathrm{d}s$$

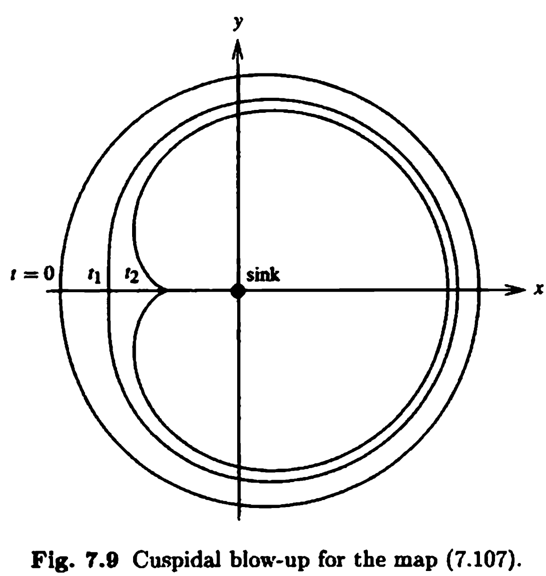


Fig. 7.9 Cuspidal blow-up for the map (7.107).

$$= -\int_{\partial\Omega} z^m \frac{\partial p}{\partial n}\,\mathrm{d}s$$
$$= \int_{|\zeta|=1} f^m \cdot \frac{1}{|\mathrm{d}f/\mathrm{d}\zeta|} \cdot \frac{Q}{2\pi}\left|\frac{\mathrm{d}f}{\mathrm{d}\zeta}\mathrm{d}\zeta\right|$$
$$= \frac{Q}{2\pi\mathrm{i}}\int_{|\zeta|=1} f^m \frac{\mathrm{d}\zeta}{\zeta};$$

this is zero for $m \geqslant 1$ because $f(0,t) = 0$. We should also mention that analytic continuation methods can sometimes be used to solve (7.106) as a functional differential equation everywhere within the unit circle; this procedure is closely related to the idea of the Schwarz function introduced in Exercise 7.18.

This discussion of blow-up leads us to the last topic we wish to mention in connection with free boundary problems.

*7.6 Regularisation

We repeat our statement that, more than any others, the partial differential equation problems discussed in this chapter raise the spectre of ill-posedness as a possible attribute of everyday models of real-world phenomena. In Chapters 4–6, we have only cited the backwards heat equation and the Cauchy problem for elliptic equations as possible examples of ill-posedness, but they are models that rarely, if ever, occur in practice.[153] As a paradigm, however, the backwards heat equation is exceedingly helpful because it is easy to analyse by, say, transform methods, and it enables us to ask the question 'suppose the modelling of a practical problem led to the backward heat equation with a small extra regularising term, say

$$\frac{\partial u}{\partial t} = -\frac{\partial^2 u}{\partial x^2} + \epsilon\mathcal{J}(u), \quad u(x,0) = u_0(x); \tag{7.108}$$

what could we say about the response as $\epsilon \to 0$?'[154]

As discussed in §6.7.1, one $\mathcal{J}$ which does indeed regularise the backward heat equation and make (7.108) into a well-posed Cauchy problem for $\epsilon > 0$ is $\mathcal{J}(u) = -\partial^4 u/\partial x^4$, in which case the Fourier transform solution is

$$u = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{u}_0(k)\mathrm{e}^{-\mathrm{i}kx+k^2t-\epsilon k^4 t}\,\mathrm{d}k. \tag{7.109}$$

As ϵ decreases to zero this integral becomes more and more irregular, tending to a function which may blow up in finite time.

Unfortunately, representations such as (7.109) are unavailable for free boundary problems and their regularisation then represents a challenging area of current research. In practice, there are always physical regularisation mechanisms available such as 'surface energy' in phase changes (or surface tension or viscosity in

[153]This statement assumes we exclude inverse problems in which we try to 'postdict' the past; they will be mentioned again in Chapter 9.

[154]This is an example of the so-called *Tikhonov regularisation* of ill-posed problems.

free surface flows), surface chemical reactions and several other sources of 'dissipation'. Indeed, these mechanisms are much more likely to exist near a free boundary, where the continuum models we have constructed have solutions that change infinitely rapidly, than away from it. The mathematical consequence is that higher derivatives can thereby be introduced into the free boundary conditions. These often have the advantage of increasing the likelihood of well-posedness, as in (7.109), but the disadvantage of being difficult to analyse.

The idea of introducing higher derivatives into free boundary conditions is not the only tool in the mathematician's regularising armoury. Indeed, we have already seen one dramatic example where a simple reformulation of the Stefan problem in terms of the enthalpy, as in (7.84), eliminated superheating. This was achieved at the expense of introducing new free boundaries and a mushy region into the solution. It is equally possible to smooth out the free boundary altogether, say by replacing $h(u)$ in (7.85) by the function in Fig. 7.10. For such a smooth monotone $h(u)$, it is possible to prove existence and uniqueness relatively easily, but the bounds that are needed cannot easily be established uniformly in the limit as $\epsilon \to 0$. Such an approach leads on to a whole hierarchy of smoothed models in which an auxiliary function $f(x,t)$ is introduced to model the fraction of material that has undergone a transition at the free boundary. Thus the Stefan problem is written as

$$\frac{\partial u}{\partial t} + L\frac{\partial f}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

together with a rate equation for the evolution of f which, in some suitable limit, would give f to be zero or unity depending which side of the free boundary was being considered. This brings us back full-circle to the ideas of reaction–diffusion equations introduced in Chapter 6.

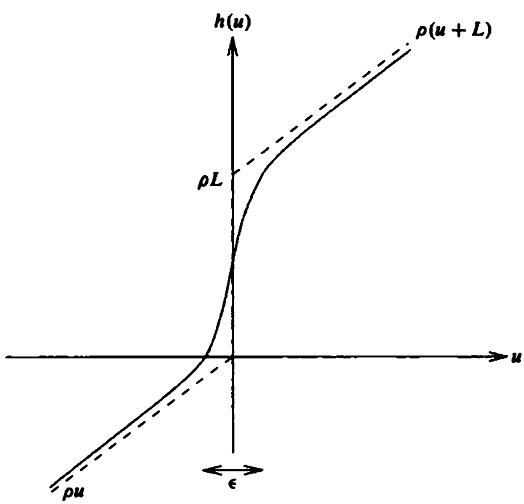


Fig. 7.10 Smoothing the enthalpy.

Finally, we remark that we could make an even more radical regularisation of the superheated or supercooled one-phase Stefan problem by identifying it with an oxygen consumption problem, as in (7.80). We have seen that this identification can be made as long as the concentration c is allowed to change sign. However, we could restrict c to be positive, thereby making the oxygen consumption model well posed, but this might entail the generation of extra components of the free boundary. Hence, if we allow the Stefan problem to admit new components in its free boundary in precisely the same way, then we are performing a regularisation by 'nucleation'.

*7.7 Postscript

We conclude this chapter with a brief introduction to a special class of free boundary problems of wide practical applicability but whose mathematical character is quite different from most of those discussed hitherto. These are problems in which the free boundary has two dimensions fewer than the dimensionality of the governing partial differential equations. Hence they can be called *codimension-two free boundary problems*, in contrast to the codimension-one problems of §7.1.

We have in fact already encountered such configurations in our discussion of contact problems in elasticity. However, contact problems are special in that the free boundary is constrained to lie in the prescribed surface in which contact occurs, and it is quite possible for a one-dimensional free boundary to move more freely in three-dimensional space. Such situations arise in modelling materials such as superfluids and superconductors, but they are most readily visualised by looking at the motion of vortices in water. We recall from §5.1.4.1 that the model for such a vortex in two-dimensional inviscid flow involves finding a velocity potential ϕ which has a singularity of the form, say,

$$\phi \sim \frac{\theta}{2\pi}$$

near a vortex at the origin of polar coordinates (r, θ). Put in terms of distributions, the velocity $\mathbf{v}$ and stream function ψ satisfy

$$\mathbf{v} = \operatorname{curl}(0, 0, \psi), \quad \operatorname{curl}\mathbf{v} = (0, 0, -\nabla^2\psi) = \big(0, 0, \delta(\mathbf{x})\big), \tag{7.110}$$

where $\mathbf{x} = (x, y)$, together with initial and boundary conditions away from the origin. This model is adequate for a vortex whose position is given *a priori*, but, if the vortex is free to move, then extra information is needed before we can formulate a free boundary model for its dynamics. In two dimensions this comes from *Helmholtz' law* which, crudely speaking, asserts that the vortex responds to the regular part of $\nabla\phi$, i.e. that a vortex whose position is $\mathbf{x}(t)$ moves such that

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \lim_{\mathbf{x}\to\mathbf{x}(t)} \nabla\left(\phi - \frac{\theta}{2\pi}\right), \tag{7.111}$$

when θ is now measured relative to $\mathbf{x}(t)$. Unfortunately, this law, which asserts that the vortex moves with the 'regular part' of the local fluid velocity, is difficult to justify without a complicated asymptotic analysis.

The situation is more interesting in three dimensions when (7.110) becomes

$$\operatorname{curl}\mathbf{v} = \boldsymbol{\delta}_\Gamma(\mathbf{x} - \mathbf{X}(s,t)),$$

where $\mathbf{x} = \mathbf{X}(s,t)$ describes the vortex curve Γ at time t, and the vector delta function is defined in terms of scalar delta functions as

$$\boldsymbol{\delta}_\Gamma(\mathbf{x}) = \int_\Gamma \delta(\mathbf{x} - \mathbf{x}')\,d\mathbf{x}'.$$

Now, referring back to (4.77), we find that

$$\mathbf{v}(\mathbf{x},t) = -\frac{1}{4\pi}\int_\Gamma \frac{(\mathbf{x} - \mathbf{x}') \wedge d\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3}, \tag{7.112}$$

and, to make further progress, we must study the behaviour of $\mathbf{v}$ as we approach Γ.[155] Unfortunately, (7.112) is singular as $\mathbf{x} \to \Gamma$ and a tedious calculation is necessary to show that, when we are a dimensionless distance d from Γ,

$$\mathbf{v} \sim \frac{1}{2\pi d}\mathbf{e}_\theta - \left(\frac{\log d}{4\pi}\right)\kappa\mathbf{b} + \cdots,$$

where $\mathbf{e}_\theta$ is a local azimuthal vector, κ is the principal curvature of Γ and $\mathbf{b}$ is the binormal. Hence, if we follow (7.111) and assert that the velocity of Γ is governed by the regular part of the locally induced flow, then the vortex velocity $\mathbf{v}_\Gamma$ is governed by the partial differential equation

$$\mathbf{v}_\Gamma = -\frac{1}{4\pi}(\log d_0)\,\kappa\mathbf{b} \tag{7.113}$$

when d_0 is some 'cut-off' that needs to be preassigned.[156] What has happened is that the intensity of the singularity at the vortex is so strong that it controls its motion independently of any externally-imposed velocity field, in sharp contrast to typical models for codimension-one problems. We note also that (7.113) is a generalisation of the equation of curvature flow (6.88) and its only known exact solution is a rotating helix given parametrically by

$$\mathbf{x} = \big(a\cos(s - \omega t), a\sin(s - \omega t), b(s - Vt)\big), \tag{7.114}$$

where a and b are constants and the constant rotation rate ω and translation speed V are related as shown in Exercise 7.21. When $\omega = 0$ and $b \to 0$ with bV finite, this solution represents a 'smoke ring', and we will return to it in §9.2.3.

[155] Had we adopted this approach for any of the codimension-one problems described earlier, say by using the integral representation (7.68), we would have found well-defined limits as we approached Γ from either side and hence the jump conditions with which we have become familiar.

[156] This device of introducing a cut-off can be thought of as a physical regularisation of the singularity that would occur if $d_0 = 0$.

Exercises

7.1. Consider the Stefan problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad x \neq s(t),$$

with

$$u = 0 \quad \text{and} \quad \left[\frac{\partial u}{\partial x}\right]_-^+ = -\dot{s}$$

at $x = s(t)$, and

$$s(0) = 0, \qquad u(x,0) = \begin{cases} u_0 = \text{constant}, & x < 0, \\ u_1 = \text{constant}, & x > 0. \end{cases}$$

Show that there is a solution $u(x,t) = U(\eta)$, where $\eta = x/\sqrt{t}$ and $s(t) = \alpha\sqrt{t}$, as long as

$$\frac{\alpha}{2}e^{\alpha^2/4} + u_0\left(\int_{-\infty}^{\alpha} e^{-\eta^2/4}\,d\eta\right)^{-1} + u_1\left(\int_{\alpha}^{\infty} e^{-\eta^2/4}\,d\eta\right)^{-1} = 0.$$

When $u_0 = 0$, use the result that $\alpha e^{\alpha^2/4}\int_\alpha^\infty e^{-\eta^2/4}\,d\eta$ is monotone and tends to 2 as $\alpha \to \infty$ to show that there is a solution for α if $u_1 > -1$ but not if $u_1 \leqslant -1$.

Remark. The Stefan condition as written here, with a latent heat of 1, forces the liquid to lie in $x > s(t)$ and the solid in $x < s(t)$. The liquid is supercooled and solidifies ($\alpha > 0$) if $u_0 = 0$ and $-1 < u_1 < 0$. Likewise, the solid is superheated and melts if $u_1 = 0$ and $0 < u_0 < 1$. The non-existence of solutions for $u_1 < -1$ or $u_0 > 1$ is discussed in §7.5.1.

7.2. Suppose, in a Muskat problem, $\nabla^2 p_w = 0$ on one side of an interface and $\nabla^2 p_o = 0$ on the other side (see §7.1.1). Also suppose that gravity is negligible and that the free boundary conditions are $\epsilon^{-1}\partial p_w/\partial n = \partial p_o/\partial n = -v_n$, where $0 < \epsilon \ll 1$. Write $p_w = \epsilon p_{w0} + \cdots$, $p_o = p_{o0} + \cdots$ and equate coefficients of ϵ to show that p_{o0} satisfies the Hele-Shaw problem

$$\nabla^2 p_{o0} = 0 \qquad \text{with} \qquad p_{o0} = 0, \quad \frac{\partial p_{o0}}{\partial n} = -v_n$$

at the free boundary. What problem does p_{w0} satisfy?

7.3. Consider the one-dimensional obstacle problem with a smooth concave obstacle $f(x)$, where f is even, $f(0) = 0$ and $d^2f/dx^2 < 0$. Consider the linear complementarity formulation of p. 331,

$$\frac{d^2u}{dx^2} \leqslant 0, \quad u - f \geqslant 0, \quad (u-f)\frac{d^2u}{dx^2} = 0,$$

with $u(-1) = u(1) = 0$. Suppose that the contact region is $-x^* < x < x^*$. Assume that, as can be justified using the theory of distributions, we can write

$$\int_{x^*-}^{x^*+} \frac{\mathrm{d}^2 u}{\mathrm{d}x^2}\,\mathrm{d}x = \left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_{x^*+} - \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x^*-}.$$

Show that, if the right-hand side of this equation is positive, then we do not have $\mathrm{d}^2u/\mathrm{d}x^2 \leqslant 0$. Show also that, if the right-hand side is negative, then we cannot have $u - f \geqslant 0$. Deduce that $\mathrm{d}u/\mathrm{d}x$ is continuous at $x = x^*$.

* 7.4. Consider an American put option as in §7.1.2, with optimal exercise value $S^*(t)$, and $V(S,t) = E - S$ for $0 \leqslant S \leqslant S^*(t)$.

(i) Suppose that at time t the option is alive and S falls to $S^*(t)$. Show that, if $\lim_{S\downarrow S^*} V(S,t) = E - S^*(t) + A$, where $A > 0$, then after the next time step $\mathrm{d}t$ the option should be exercised with probability $1/2 + O(\mathrm{d}t)$ for a profit of $A + O(\sqrt{\mathrm{d}t})$, while with probability $1/2 + O(\mathrm{d}t)$ its value will change by only $O(\sqrt{\mathrm{d}t})$. Deduce that arbitrage forces V to be continuous at $S = S^*$.

(ii) Now suppose that $S = S^*$ as in part (i) and that $\partial V/\partial S\,(S^*,t) < -1$. Show that the option value falls below the payoff for values of S just above S^* and explain why this is impossible. Finally, suppose that $\partial V/\partial S\,(S^*,t) > -1$, and show by a sketch of V as a function of S that the option value would be greater if S^* were smaller, and that taking S^* smaller would decrease the value of $\partial V/\partial S\,(S^*,t)$. Deduce that the option has its maximum value to the holder if $S^*(t)$ is chosen to make $\partial V/\partial S$ continuous there.

7.5. The linear complementarity form of the American put option problem of p. 334 is

$$\mathcal{L}V = \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV \leqslant 0, \quad V - \max(E - S, 0) \geqslant 0,$$

$$(\mathcal{L}V)\,(V - \max(E - S, 0)) = 0,$$

together with $V(S,T) = \max(E - S, 0)$. Adapt the argument of Exercise 7.3 to show that $\partial V/\partial S$ is continuous at the optimal exercise boundary $S = S^*(t)$.

* 7.6. Slow viscous flow in two dimensions is modelled by

$$\mu\nabla^2\mathbf{u} = \nabla p, \quad \nabla\cdot\mathbf{u} = 0,$$

where $\mu =$ constant, $\mathbf{u} = (u, v)$ and the force per unit area on a surface with normal $\mathbf{n}$ is

$$\begin{pmatrix} -p + 2\mu\,\partial u/\partial x & \mu\,(\partial u/\partial y + \partial v/\partial x) \\ \mu\,(\partial u/\partial y + \partial v/\partial x) & -p + 2\mu\,\partial v/\partial y \end{pmatrix}\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}.$$

Show that

(i) there is a ψ such that

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x},$$

and that

$$\nabla^4 \psi = 0;$$

(ii) there is an A such that

$$-p + 2\mu \frac{\partial u}{\partial x} = -2\mu \frac{\partial^2 A}{\partial y^2}, \qquad -p + 2\mu \frac{\partial v}{\partial y} = -2\mu \frac{\partial^2 A}{\partial x^2},$$

$$\mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) = 2\mu \frac{\partial^2 A}{\partial x \, \partial y},$$

and that

$$\nabla^4 A = 0.$$

In a steady slow flow with a stress-free free boundary Γ, show that $\psi =$ constant and $\partial/\partial s(\nabla A) = \mathbf{0}$ on Γ, where $\partial/\partial s$ denotes the derivative along Γ. Show that this implies $A =$ constant and $\partial A/\partial n = 0$ on Γ. In what geometries can A be set equal to zero on Γ without loss of generality?

* 7.7. Suppose there is a constant surface tension in the slow flow of Exercise 7.6 so that the force per unit area on the free surface is $-T\kappa\mathbf{n}$, where $T =$ constant and κ is the curvature. Show that, without loss of generality,

$$A = \text{constant}, \quad \frac{\partial A}{\partial n} = \frac{T}{2\mu}.$$

7.8. Suppose a smooth rigid indenter $y = f(x)$ displaces a smooth elastic half-space by a small amount over the region $|x| < c$. Denote the force per unit area on the boundary of the half-space by

$$\begin{pmatrix} \sigma_x & \tau \\ \tau & \sigma_y \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Formulate boundary conditions for the displacement (u, v) and forces on $y = 0$ as

$$\sigma_y = \tau = 0, \quad v < f(x) \qquad \text{for } |x| > c,$$
$$\tau = 0, \quad v = f(x), \quad \sigma_y < 0 \quad \text{for } |x| < c.$$

Write these conditions in linear complementarity form.

7.9. Suppose that

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } 0 < x < s(t),$$

with

$$u = 0, \quad \frac{\partial u}{\partial x} = -\dot{s} \quad \text{at } x = s(t),$$
$$\frac{\partial u}{\partial x}(0,t) = 0 \quad \text{and} \quad u(x,0) = u_0(x) \leqslant 0 \qquad \text{with} \qquad s(0) = s_0.$$

Show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}s^2 + \int_0^{s(t)} xu(x,t)\,\mathrm{d}x\right) = u(0,t) \leqslant 0,$$

and hence that, if

$$\frac{1}{2}s_0^2 + \int_0^{s_0} xu_0(x)\,\mathrm{d}x < 0,$$

then neither can there be a steady state in which $u = 0$ and $s > 0$, nor can s vanish.

Remark. This method of proving blow-up can be extended to functions

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(f(s) + \int_0^{s} f'(x)u(x,t)\,\mathrm{d}x\right)$$

as long as f has suitable properties.

$*$7.10. Show that

$$\lim_{x\to 0}\int_0^\infty \mathrm{e}^{-(x-t)^2/4t}\frac{\mathrm{d}t}{\sqrt{t}} = \int_0^\infty \mathrm{e}^{-t/4}\frac{\mathrm{d}t}{\sqrt{t}},$$

and that

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_0^\infty \mathrm{e}^{-(x-t)^2/4t}\frac{\mathrm{d}t}{\sqrt{t}} = -\frac{1}{2}\int_0^\infty (x-t)\mathrm{e}^{-(x-t)^2/4t}\frac{\mathrm{d}t}{t^{3/2}} \quad \text{for } x > 0.$$

However, show also that this last derivative at $x = 0$ is *not* equal to

$$\frac{1}{2}\int_0^\infty \mathrm{e}^{-t/4}\frac{\mathrm{d}t}{\sqrt{t}}.$$

Remark. It can be shown that

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_0^\infty f(t)\mathrm{e}^{-(x-t)^2/4t}\frac{\mathrm{d}t}{\sqrt{t}}\bigg|_{x=0} = \frac{1}{2}\int_0^\infty f(t)\mathrm{e}^{-t/4}\frac{\mathrm{d}t}{\sqrt{t}} - \sqrt{\pi}f(0),$$

which explains the factor 1/2 on the left-hand side of (7.68).

7.11. A circular membrane has zero transverse displacement on the circle $x^2 + y^2 = 1$, $z = 0$, but it is constrained to lie above the smooth obstacle $z = \epsilon(1 - 2(x^2 + y^2))$, where ϵ is small and positive. Show that the radius R of the contact region satisfies $2R^2(1 - 2\log R) = 1$.

7.12. The conditions ahead of (0) and behind (1) a gasdynamic shock moving with speed V_s are related by (2.49), namely

$$\begin{aligned}[\rho(V_s-u)]_0^1 &= [p+\rho(V_s-u)^2]_0^1 \\ &= \left[\rho(V_s-u)\left(\frac{\gamma p}{(\gamma-1)\rho}+\frac{1}{2}(V_s-u)^2\right)\right]_0^1 \\ &= 0.\end{aligned}$$

Deduce that, if a piston moves with speed V_p into a tube of stationary gas in which $p = p_0$ and $\rho = \rho_0$, then the free boundary (shock) has speed V_s, where

$$V_s^2 - \frac{\gamma+1}{2}V_pV_s - a_0^2 = 0, \quad a_0^2 = \frac{\gamma p_0}{\rho_0}.$$

$*$7.13. Show that the Rankine–Hugoniot conditions of Exercise 7.12 imply

$$(u_1-u_0)^2 = (p_1-p_0)\left(\frac{1}{\rho_0}-\frac{1}{\rho_1}\right) \quad \text{and} \quad \frac{\rho_0}{\rho_1} = \frac{(\gamma+1)p_0+(\gamma-1)p_1}{(\gamma-1)p_0+(\gamma+1)p_1}.$$

Deduce that, if a shock wave with pressure p_1 behind it propagating into stationary gas in which $p = p_0$ and $\rho = \rho_0$ reflects from a plane wall to which it is parallel, then the pressure behind the reflected shock is p_2, where

$$\frac{p_2}{p_1} = \left((3\gamma-1)\frac{p_1}{p_0}-(\gamma-1)\right)\Big/\left((\gamma-1)\frac{p_1}{p_0}+(\gamma+1)\right).$$

$*$7.14. A two-dimensional jet of inviscid irrotational fluid, of thickness $2h_\infty$ and moving to the right with speed 1, enters a semi-infinite rectangular cavity with walls at $y = \pm 1$, as shown in Fig. 7.11; the y axis is tangent to the free surface.

Ignoring gravity and surface tension, show that the boundary value problem for the complex potential $w(z) = \phi + i\psi$ for the upper half of the flow is that $w(z)$ is analytic in the fluid region, with

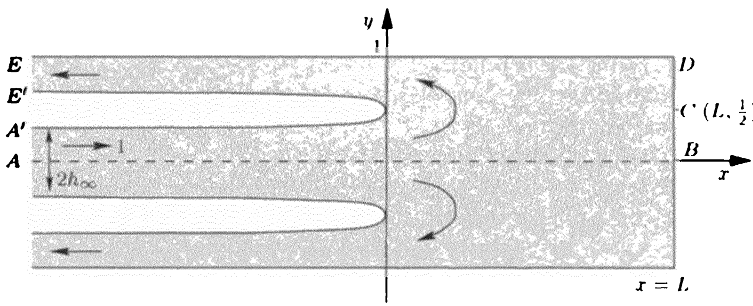


Fig. 7.11 A jet entering a box.

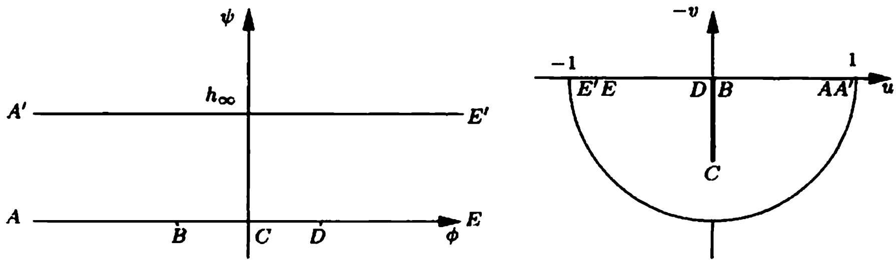


Fig. 7.12 Potential and hodograph planes for the flow of Exercise 7.14.

$$\psi = 0 \quad \text{on } ABCDE,$$
$$\psi = h_\infty, \quad |\nabla\phi|^2 = 1 \quad \text{on } E'A'.$$

Taking the reference point for ϕ so that $w = 0$ at C, show that the potential and hodograph ($u - iv = dw/dz$) planes are as in Fig. 7.12. When these two planes are mapped onto each other (do not attempt this unless you are feeling very strong), to give $dw/dz = F(w)$ for some holomorphic F, the solution of this differential equation contains one arbitrary constant. Noting that the positions of both A (or, by symmetry, E) and B are specified, deduce that there is a relation between h_∞ and L.
Now consider the case $L = \infty$, with stagnant fluid far inside the cavity. Show that B, C and D coincide at the origin in the hodograph plane, and that the flow domain is the whole interior of the semicircle shown. Show that

$$\frac{dw}{dz} = \frac{1 - e^{\pi w/2h_\infty}}{1 + e^{\pi w/2h_\infty}} = -\tanh\frac{\pi w}{4h_\infty}.$$

Find w satisfying $w = ih_\infty$ at $z = i/2$, the tip of the air finger shown in Fig. 7.11. Show that the free surface for this flow, $w = \phi + ih_\infty$, $-\infty < \phi < \infty$, satisfies

$$e^{\pi x/2h_\infty} \cos\frac{\pi(y - \frac{1}{2})}{2h_\infty} = 1.$$

Finally, show that the condition $y \to \pm h_\infty$ as $x \to -\infty$ is only consistent with this equation if $h_\infty = \frac{1}{4}$, so that the finger occupies half of the cavity, and that the free surface is the same shape as the Grim Reaper of Exercise 6.33 and the Saffman–Taylor finger of Exercise 7.19.

Remark. It can be shown that h_∞ is an increasing function of L. As $L \to 0$, the flow consists of a thin jet along AB which turns through a right angle at B (this flow can easily be analysed in its own right), runs up BCD, turns through another right angle at D and finally runs along DE.

* 7.15. Inviscid fluid flows in the x direction with unit speed past a plate $x = 0$, $|y| < 1$. There is a wake, bounded by the separation streamlines $y = \pm f(x)$, $x > 0$,

in which the pressure is constant. Show that the free boundary problem for the separation streamline $y = f(x)$, $x > 0$, is

$$\nabla^2\phi = 0 \quad \text{for } y > 0, \qquad \phi \sim x \quad \text{at infinity,}$$

with

$$\frac{\partial \phi}{\partial y}(x,0) = 0 \quad \text{for } x < 0$$

and

$$\frac{\partial \phi}{\partial n} = 0, \quad |\nabla\phi| = 1 \quad \text{on } y = f(x),\ x > 0.$$

Show that the flow region in the plane of $w = \phi + \mathrm{i}\psi$ can be taken to be $\psi \geqslant 0$, with $\psi = 0$, $\phi < 0$ being the upstream dividing streamline, $\psi = 0$, $0 < \phi < \phi_0$ the plate and $\psi = 0$, $\phi_0 < \phi$ the free boundary. Show also that, in the plane of $W = \log(\mathrm{d}w/\mathrm{d}z)$, these three curves are $\Im W = 0$, $0 > \Re W > -\infty$; $\Im W = -\pi/2$, $-\infty < \Re W < 0$; $\Re W = 0$, $-\pi/2 < \Im W < 0$, respectively. The Schwarz–Christoffel result can be used to show that the flow region in the W plane can be related to that in the w plane (namely $\Im w > 0$) uniquely by the formula

$$w = \phi_0 \operatorname{cosech}^2 W, \quad \text{where } \phi_0 = \phi(0,1).$$

Show that this leads to the holomorphic differential equation

$$\frac{\mathrm{d}w}{\mathrm{d}z} = \sqrt{1 - \frac{\phi_0}{w}} - \sqrt{-\frac{\phi_0}{w}}.$$

Integrate this equation along the plate to obtain an equation for ϕ_0, and then integrate along the free boundary to show that it is given parametrically by

$$z = \mathrm{i}\left(1 + \int_{\phi_0}^{\phi} \frac{\sqrt{w}\,\mathrm{d}w}{\sqrt{\phi_0 - w} - \sqrt{\phi_0}}\right).$$

Remark. The Schwarz–Christawful formula says that, if α_i are the interior angles of a closed polygon, then the map

$$z - z_0 = \kappa f(\zeta), \quad \kappa,\ z_0 \text{ complex constants,}$$

where

$$\frac{\mathrm{d}f}{\mathrm{d}\zeta} = \prod_1^n (\zeta - \zeta_i)^{\alpha_i/\pi - 1}, \quad \zeta_i \text{ real constants,}$$

takes the real axis in the ζ plane into a polygon with these interior angles, with ζ_i mapping into the vertices of the polygon and the upper half of the ζ plane mapping into the interior of the polygon. For fixed ζ_i, different choices of z_0, $\arg\kappa$ and $|\kappa|$ correspond to translation of the polygon, rotation of the polygon, and expansion or contraction of the polygon with the ratio of its sides fixed, respectively. Hence, given a target polygon with vertices z_i, the images of any three points ζ_1, ζ_2 and ζ_3 can be chosen to be z_1, z_2 and z_3, respectively; however, the shape of the polygon is then uniquely determined by $\zeta_4, \zeta_5, \dots$ and the Riemann mapping theorem shows that $z_4, z_5, \dots$ do indeed define ζ_4, ζ_5, etc. uniquely.

*7.16. At time $t = 0$, a two-dimensional body $y = f(x) - Vt$ (where $f(0) = 0$, $\mathrm{d}^2 f/\mathrm{d}x^2 \geqslant 0$ and f is even in x) impacts a half-space of inviscid liquid $y < 0$ from above. The free boundary problem for the velocity potential ϕ is, in the absence of gravity,

$$\nabla^2 \phi = 0 \quad \text{in the liquid, with} \quad \phi \to 0 \quad \text{as } y \to -\infty,$$

with

$$\frac{\partial \phi}{\partial y} = \frac{\partial \eta}{\partial t} + \frac{\partial \phi}{\partial x}\frac{\partial y}{\partial x}, \quad \frac{1}{2}|\nabla \phi|^2 + \frac{\partial \phi}{\partial t} = 0$$

on the free boundary $y = \eta(x,t)$, and

$$\frac{\partial \phi}{\partial y} = -V + \frac{\mathrm{d}f}{\mathrm{d}x}\frac{\partial \phi}{\partial x}$$

on the wetted surface of the body.

Show that V can be taken equal to unity without loss of generality and that, when the body is a wedge, with $f = \alpha|x|$, the problem is then invariant under the transformations $x' = \mathrm{e}^{\lambda}x$, $y' = \mathrm{e}^{\lambda}y$, $t' = \mathrm{e}^{\lambda}t$, $\eta' = \mathrm{e}^{\lambda}\eta$, $\phi' = \mathrm{e}^{\lambda}\phi$. Deduce the existence of a similarity solution $\phi = t\Phi(X,Y)$, $\eta(x,t) = tH(X)$, $x/t = X$, $y/t = Y$, where

$$\frac{\partial^2 \Phi}{\partial X^2} + \frac{\partial^2 \Phi}{\partial Y^2} = 0 \quad \text{in the liquid,}$$

with

$$\Phi - \left(X\frac{\partial \Phi}{\partial X} + Y\frac{\partial \Phi}{\partial Y}\right) + \frac{1}{2}\left(\left(\frac{\partial \Phi}{\partial X}\right)^2 + \left(\frac{\partial \Phi}{\partial Y}\right)^2\right) = 0,$$

$$\frac{\partial \Phi}{\partial Y} = H - \frac{\mathrm{d}H}{\mathrm{d}X} + \frac{\partial \Phi}{\partial X}\frac{\mathrm{d}H}{\mathrm{d}X}$$

on $Y = H(X)$, and

$$\frac{\partial \Phi}{\partial Y} = -1 + \alpha\frac{\partial \Phi}{\partial X} \quad \text{on the wetted part of } Y = \alpha X - 1,$$

$$\frac{\partial \Phi}{\partial Y} = 0 \quad \text{on } X = 0, Y < -1.$$

*7.17. Suppose that the impacting body in Exercise 7.16 is $y = \epsilon\,(f(x) - t)$, where $0 < \epsilon \ll 1$ and t is not too large. Show that, if $\phi = \epsilon\phi_0 + \cdots$, $\eta = \epsilon\eta_0 + \cdots$

and terms quadratic in ϵ are neglected, then the boundary conditions can all be imposed on $y = 0$ in the form

$$\frac{\partial \phi_0}{\partial t} = 0, \quad \frac{\partial \eta_0}{\partial t} = \frac{\partial \phi_0}{\partial y} \quad \text{for } |x| > d(t),$$

and

$$\frac{\partial \phi_0}{\partial y} = -1 \quad \text{for } |x| < d(t),$$

for some function $d(t)$. Taking $\phi_0 = 0$ on $y = 0$, $|x| > d(t)$, use the methods of §5.9 to show that

$$\phi_0 = \Im\left(-z + \sqrt{z^2 - d^2}\right)$$

is the solution with the least singular behaviour at the 'codimension-two' free boundary $|x| = \pm d(t)$. Deduce that the pressure, which from Bernoulli's equation is approximately $-\partial\phi_0/\partial t$, is infinite at these points.

Remark. The evolution of $d(t)$ can be predicted in terms of $f(x)$ if it is assumed that $\eta_0(d(t) + 0, t) = f(d(t)) - t$.

∗7.18. Suppose that p satisfies a Hele-Shaw free boundary problem in which $\nabla^2 p = 0$, with $p = 0$ and $\partial p/\partial n = -v_n$ at the free boundary, denoted by $t = \omega(x, y)$. Show that

$$u(x, y, t) = \int_\omega^t p(x, y, \tau)\, d\tau$$

satisfies the obstacle problem

$$\nabla^2 u = 1 \qquad \text{with} \qquad u = \frac{\partial u}{\partial n} = 0 \quad \text{on } t = \omega.$$

Show also that

$$g(z, t) = \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\left(u - \frac{1}{4}(x^2 + y^2)\right)$$

is analytic, and that $\bar{z} = g(z, t)$ on the free boundary.

Remark 1. An analytic curve $f(x, y) = 0$ can be written as $\bar{z} = g(z)$, where $f\left((z + g(z))/2, (z - g(z))/2i\right) = 0$. The function g is called the *Schwarz function* of the curve and you may like to show that it satisfies $g(\bar{g}(z)) = z$, which is the consistency condition necessary when we replace one real equation ($f(x, y) = 0$) by the complex equation $\bar{z} = g(z)$. Its determination involves the solution of a Cauchy problem for its real or imaginary parts, and so it is very likely to have singularities close to the curve.

Remark 2. For the Hele-Shaw problem above, the singularities of g within the fluid are independent of t unless they coincide with those of p, because $\partial u/\partial t = p$.

∗7.19. Consider Hele-Shaw flow in a parallel-sided channel $-\pi < y < \pi$, $-\infty < x < \infty$, in which the fluid is removed at a constant rate from $x = +\infty$

so that its velocity there is $(V, 0)$. Show that, if $z = f(\zeta, t)$ maps the fluid region onto $|\zeta| < 1$ and the sink at $x = +\infty$ is mapped onto $\zeta = 0$, then $f(\zeta, t) = -\log\zeta + O(1)$ as $\zeta \to 0$, and that Q in (7.106) is replaced by $-2\pi V$. Use (7.106) to show that the map

$$z = f(\zeta, t) = \frac{Vt}{\lambda} - \log\zeta + 2(1-\lambda)\log\left(\frac{1+\zeta}{2}\right)$$

gives a travelling wave solution, moving with speed $U = V/\lambda$, for any value of λ between 0 and 1. Set $\zeta = e^{i\theta}$ to show that the interface has the equation $e^{(x-Ut)/2(1-\lambda)} = \cos(y/2\lambda)$ and hence show that as $x \to -\infty$ the interface is asymptotic to the lines $y = \pm\lambda\pi$.

Remark. There is an exact unsteady solution in which a small, nearly sinusoidal perturbation to a planar interface evolves into this *Saffman–Taylor finger* as $t \to \infty$. The value of λ is not specified by the simple model we have used, and the question of its determination is an example of 'pattern selection'. Note that the value $\lambda = 1/2$, which is often seen in experiments, gives the Grim Reaper solution of curvature flow (see Exercise 6.33).

7.20. When gravity is important in a Muskat problem, the free boundary conditions are

$$p_w = p_o, \quad -K_w\frac{\partial}{\partial n}(p_w + \rho_w gy) = -K_o\frac{\partial}{\partial n}(p_o + \rho_o gy) = v_n,$$

where p_w and p_o are the pressures on either side of the boundary, $K_{w,o} = k_{w,o}/\mu_{w,o}$ and y is vertical. Show that a travelling wave solution exists in which the free boundary is $y = Vt$ and

$$p_w = -\left(\frac{V}{K_w} + \rho_w g\right)(y - Vt), \quad p_o = -\left(\frac{V}{K_o} + \rho_o g\right)(y - Vt).$$

Show that, if the free boundary is perturbed to $y = Vt + \epsilon e^{\sigma t}\sin kx$, $k > 0$, and ϵ is small, then

$$\frac{\sigma}{k} = \frac{gK_wK_o(\rho_o - \rho_w) + (K_w - K_o)V}{K_w + K_o},$$

and deduce that, even if a heavy and more viscous fluid overlies a light one, the flow is stable if V is sufficiently large and negative.

7.21. Show that the principal normal $\mathbf{n}$ and binormal $\mathbf{b}$ of (7.114) are

$$\mathbf{n} = (-\cos(s - \omega t), -\sin(s - \omega t), 0),$$

$$\mathbf{b} = \frac{1}{\sqrt{a^2 + b^2}}(b\sin(s - \omega t), -b\cos(s - \omega t), a),$$

and that the curvature is $a/(a^2 + b^2)$. Deduce that the velocity of a point on (7.114) satisfies (7.113) as long as

$$V = \omega + \frac{1}{4\pi}\frac{\log d_0}{b\sqrt{a^2 + b^2}}.$$

8

Non-quasilinear equations

8.1 Introduction

Several practical situations give rise to partial differential equations that are not quasilinear. For example, suppose $u(x, y)$ is the elevation of a pile of dry sand that is heaped on a table in limiting equilibrium, so that if any more sand was poured from above it would slide off. Then the angle between the normal to the surface $u = u(x, y)$ and the vertical $(0, 0, 1)$ is a prescribed constant γ, the angle of friction, so that

$$\left(1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right)^{-1/2} = \cos\gamma.$$

With a slight change of variables, this can be transformed into the famous *eikonal equation*

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 = 1. \tag{8.1}$$

There is another application of (8.1) that has implications far beyond those for sand piles. This concerns approximate solutions of Helmholtz' equation describing wave propagation in the frequency domain. We recall that the equation was derived from the wave equation $\partial^2\phi/\partial\tau^2 = a_0^2\nabla^2\phi$ by writing

$$\phi = \Re\left(e^{-i\omega\tau}\psi(x, y)\right), \tag{8.2}$$

where we have replaced t in the wave equation by τ to avoid later confusion in notation. In dimensionless variables this gives Helmholtz' equation

$$(\nabla^2 + k^2)\psi = 0, \tag{8.3}$$

as in (5.27), where $k = \omega L/a_0$ and L is a characteristic length scale over which the waves propagate. Now, for the propagation of light, $\omega/a_0 \sim 10^7\,\mathrm{m}^{-1}$ typically, and so, on an everyday length scale for which $L \sim 1\,\mathrm{m}$, k is large. In this case, the asymptotic approximation of Helmholtz' equation is called the theory of *geometric optics*. It relies on the *WKB* asymptotic procedure for ordinary differential equations, which suggests that we write

$$\psi \sim A(x, y; k)e^{iku(x,y)} \quad \text{as } k \to \infty, \tag{8.4}$$

so that

$$\nabla\psi \sim (\mathrm{i}kA\nabla u + \nabla A)\mathrm{e}^{\mathrm{i}ku},$$

$$\nabla^2\psi \sim (-k^2A|\nabla u|^2 + 2\mathrm{i}k\nabla A\cdot\nabla u + \mathrm{i}kA\nabla^2 u + \nabla^2 A)\mathrm{e}^{\mathrm{i}ku},$$

and equating terms of $O(k^2)$ leads at once to (8.1) for u.[157]

Equation (8.1) also has its origins in the less well known, but no less interesting, problem of heat conduction over short time intervals. We have seen in (6.35) that, in an infinite two-dimensional medium, the response to a heat source at $x = y = t = 0$ is the Green's function $(1/4\pi t)\mathrm{e}^{-(x^2+y^2)/4t}$. Thus, when there are boundaries in the problem, we are motivated to try the WKB ansatz

$$T \sim \frac{1}{4\pi t}\mathrm{e}^{-V(x,y)/t}$$

in the heat conduction equation $\partial T/\partial t = \nabla^2 T$ to give, to lowest order as $t \to 0$,

$$\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 = V.$$

Thus $u = 2\sqrt{V}$ satisfies (8.1).

In the light of the models above, we expect the solutions of (8.1) to have interesting properties, since they should predict geometrical shapes ranging from the topography of heaps of granular materials to the patterns of light rays. Certainly, non-quasilinear ordinary differential equations often have properties that are much more interesting than quasilinear ones, as demonstrated in [32].

In one sense, however, there is no need to devote a chapter to this topic because, as explained in Chapter 2, *any* partial differential equation can always be written as a quasilinear system. However, this can often only be done at the expense of cross-differentiation, which leads to loss of information and a greatly expanded system; also, from the practical point of view, it is much easier to interpret mathematical predictions in terms of the basic physical variables rather than complicated derivatives thereof.

8.2 Scalar first-order equations

8.2.1 Two independent variables

We recall from §1.3 that the quasilinear first-order equation

$$a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} = c \tag{8.5}$$

could be approached by exploiting either its geometrical interpretation or the ease with which all the derivatives of u can be computed in terms of known Cauchy

[157]Those familiar with the 'stationary phase' approximation to Fourier integrals will be able to show that, when a Fourier transform solution of (8.3) is available, its high-frequency limit is often dominated by the contribution from a saddle point whose location is found by a procedure precisely equivalent to the solution of (8.1) by Charpit's method, as described in the next section.

data. Neither of these options is immediately open when we study equations of the form

$$F(x, y, u, p, q) = 0, \tag{8.6}$$

where $p = \partial u/\partial x$ and $q = \partial u/\partial y$. Moreover, our earlier theory suggests that there is little chance of reducing (8.6) to a system of ordinary differential equations as we could in the quasilinear case. We say this because cross-differentiation leads to the system

$$\frac{\partial F}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial F}{\partial q}\frac{\partial q}{\partial x} = -\frac{\partial F}{\partial x} - p\frac{\partial F}{\partial u}, \tag{8.7}$$

$$\frac{\partial F}{\partial p}\frac{\partial p}{\partial y} + \frac{\partial F}{\partial q}\frac{\partial q}{\partial y} = -\frac{\partial F}{\partial y} - q\frac{\partial F}{\partial u}, \tag{8.8}$$

$$\frac{\partial u}{\partial x} = p, \tag{8.9}$$

$$\frac{\partial u}{\partial y} = q, \tag{8.10}$$

$$\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x} = 0 \tag{8.11}$$

for the vector $\mathbf{w} = (u, p, q)^{\mathsf{T}}$. Hence, when we choose three suitably independent equations from this set, as will be done below, we expect to be in the situation described in §1.3; only very rarely would a system of partial differential equations with three scalar dependent variables be integrable along characteristics. However, (8.7)–(8.11) is a very special system, as we will see shortly.

From the geometrical point of view, (8.6) is much more complicated than (8.5). It says that the normal to the solution surface at each point must lie in a cone (called the *normal cone*, despite possible confusion with §2.6) since, when we 'freeze' x, y and u, (8.6) is a relation just between the direction cosines of the normal to the solution surface $u = u(x, y)$. Only in the quasilinear case does the normal cone degenerate into a plane.

Let us now address the Cauchy problem of determining $\mathbf{w}$ from a knowledge of its values on an arbitrary curve in the (x, y) plane. Using t to parametrise the characteristics, we soon find that the partial derivatives of $\mathbf{w}$ must also satisfy the relations

$$\dot{x}\frac{\partial u}{\partial x} + \dot{y}\frac{\partial u}{\partial y} = \dot{u}, \tag{8.12}$$

$$\dot{x}\frac{\partial p}{\partial x} + \dot{y}\frac{\partial p}{\partial y} = \dot{p}, \tag{8.13}$$

$$\dot{x}\frac{\partial q}{\partial x} + \dot{y}\frac{\partial q}{\partial y} = \dot{q}, \tag{8.14}$$

where, as usual, $\dot{} = d/dt$. We can discard (8.9), (8.10) and (8.12) for the time being because, together with (8.6), they are consistent equations for $\partial u/\partial x$ and

$\partial u/\partial y$. Furthermore, by eliminating $\partial p/\partial x$ and $\partial q/\partial y$, the remaining equations can be written as

$$\dot{y}\frac{\partial F}{\partial p}\frac{\partial p}{\partial y} - \dot{x}\frac{\partial F}{\partial q}\frac{\partial q}{\partial x} = -\dot{y}\left(\frac{\partial F}{\partial y} + q\frac{\partial F}{\partial u}\right) - \dot{q}\frac{\partial F}{\partial q}, \tag{8.15}$$

$$-\dot{y}\frac{\partial F}{\partial p}\frac{\partial p}{\partial y} + \dot{x}\frac{\partial F}{\partial q}\frac{\partial q}{\partial x} = -\dot{x}\left(\frac{\partial F}{\partial x} + p\frac{\partial F}{\partial u}\right) - \dot{p}\frac{\partial F}{\partial p}, \tag{8.16}$$

$$\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x} = 0. \tag{8.17}$$

The characteristics are, as usual, defined as the curves on which the normal derivative of **w** is not determined uniquely by (8.6) and its value on the curve; hence, on a characteristic,

$$\dot{y}\frac{\partial F}{\partial p} - \dot{x}\frac{\partial F}{\partial q} = 0. \tag{8.18}$$

Moreover, consistency in (8.15)–(8.17) requires that the right-hand sides of (8.15) and (8.16) be equal and opposite and hence zero, by (8.18).

Lastly, we have freedom to choose our parameter t, and easily the most convenient choice is to make

$$\dot{x} = \frac{\partial F}{\partial p},$$

so that

$$\dot{y} = \frac{\partial F}{\partial q}.$$

The consistency conditions now reduce to

$$\dot{p} = -\frac{\partial F}{\partial x} - p\frac{\partial F}{\partial u}, \quad \dot{q} = -\frac{\partial F}{\partial y} - q\frac{\partial F}{\partial u},$$

and finally we return to (8.12) to see that

$$\dot{u} = p\frac{\partial F}{\partial p} + q\frac{\partial F}{\partial q}.$$

Gathering these equations together, we have

$$\begin{aligned} \dot{x} &= \frac{\partial F}{\partial p}, \quad \dot{y} = \frac{\partial F}{\partial q}, \quad \dot{u} = p\frac{\partial F}{\partial p} + q\frac{\partial F}{\partial q}, \\ \dot{p} &= -\frac{\partial F}{\partial x} - p\frac{\partial F}{\partial u}, \quad \dot{q} = -\frac{\partial F}{\partial y} - q\frac{\partial F}{\partial u}, \end{aligned} \tag{8.19}$$

and these are called *Charpit's equations*. Given appropriate initial data at $t = 0$, they have a unique solution, at least for small t, but, whereas this initial data was readily available in the quasilinear case, it now demands knowledge of p and q as well as u at the initial curve $t = 0$. To be precise, given Cauchy data $x = x_0(s)$, $y = y_0(s)$ and $u = u_0(s)$, we need to be able to solve the two equations

$$F(x_0, y_0, u_0, p_0, q_0) = 0 \quad \text{and} \quad u_0' = p_0 x_0' + q_0 y_0'$$

simultaneously for p_0 and q_0, where $' = \mathrm{d}/\mathrm{d}s$, and the non-vanishing of $y'\,\partial F/\partial p - x'\,\partial F/\partial q$ is the condition for local solvability, as already revealed in (8.18). Then, if we write the corresponding solution of (8.19) as

$$x = x(s,t), \quad y = y(s,t), \quad u = u(s,t), \quad p = p(s,t), \quad q = q(s,t),$$

and if $|\partial(x,y)/\partial(s,t)|$ is neither zero nor infinite, we assert that

- the result of eliminating s and t gives the solution $u = u(x,y)$ of the partial differential equation $F(x,y,u,p,q) = 0$.

If this statement is true, it is a spectacular result because we have reduced the task of solving an arbitrary scalar first-order partial differential equation with two independent variables to that of solving five autonomous ordinary differential equations, something which, as mentioned above, we have no right to expect. We must clearly scrutinise the statement much more closely but, before we do that, we issue two warnings.

Firstly, the determination of p_0 and q_0 is almost certainly non-unique unless the original equation (8.6) is quasilinear. Hence, as with non-quasilinear ordinary differential equations, extra information is usually needed before we can begin our integration of Charpit's equations.

Secondly, since the ordinary differential equations (8.19) are nonlinear, it is quite likely that their solution blows up at some finite time or, even if it exists globally, it may possibly behave chaotically. The same was true for the quasilinear case but now, when we consider the global structure of the solution, we may expect singularities other than 'shocks' to be capable of developing.

Let us now return to our assertion above. Its justification requires not only that we verify that u explicitly satisfies (8.6), but also that $\mathbf{w}$ is such that $p = \partial u/\partial x$ and $q = \partial u/\partial y$, since these relations are not guaranteed by the assertion (remember that p and q were originally introduced by these relations, but from (8.7) onwards we have been thinking of $\mathbf{w}$ as a vector with three independent components). The verification that $\mathbf{w}$ satisfies (8.6) is immediate since

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial x}\dot{x} + \frac{\partial F}{\partial y}\dot{y} + \frac{\partial F}{\partial u}\dot{u} + \frac{\partial F}{\partial p}\dot{p} + \frac{\partial F}{\partial q}\dot{q} = 0$$

on using (8.12)-(8.14); at $t = 0$, $F = 0$, so that $F \equiv 0$ for all t and s, and hence for all x and y.

To show that $p = \partial u/\partial x$ and $q = \partial u/\partial y$, it is sufficient to show that $\dot{u} = p\dot{x} + q\dot{y}$ and $u' = px' + qy'$. The former condition is immediate from (8.12)-(8.14); for the latter, we consider the function $\phi(s,t) = u' - px' - qy'$. Then, differentiating ϕ and using the partial derivative of $\dot{u} = p\dot{x} + q\dot{y}$ with respect to s, we have

$$\begin{aligned}\dot{\phi} &= p'\dot{x} + q'\dot{y} - \dot{p}x' - \dot{q}y' \\ &= \frac{\partial F}{\partial x}x' + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial p}p' + \frac{\partial F}{\partial q}q' + \frac{\partial F}{\partial u}(px' + qy') \\ &= \frac{\partial F}{\partial s} - \phi\frac{\partial F}{\partial u},\end{aligned}$$

where $\partial F/\partial s$ is calculated keeping t constant. But $F \equiv 0$, so that $\partial F/\partial s = 0$; also $\phi = 0$ at $t = 0$. Hence $\phi = 0$ for all t (no matter what the function $\partial F/\partial u$ is) and $u' = px' + qy'$, so that p and q are indeed the derivatives of u.

We can now generalise our definition of characteristics by saying that Charpit's equations (8.19) define characteristics in the five-dimensional space (x, y, u, p, q), or, equivalently, a *characteristic strip* in the three-dimensional (x, y, u) space. That is, at each point (x, y) a surface element is defined, and the solution surface is formed by 'glueing together' the characteristic strips. This solution only exists where the Jacobian $|\partial(x, y)/\partial(s, t)| \neq 0$, that is when

$$x'\frac{\partial F}{\partial q} \neq y'\frac{\partial F}{\partial p}, \tag{8.20}$$

and, as already mentioned, this condition must be satisfied by the Cauchy data. (A simpler derivation of (8.14)–(8.20) for the case $\partial F/\partial u = 1$ is given in [5] and Exercise 8.5.)

A geometrical derivation of Charpit's equations follows from the observation that the *dual cone* of the normal cone, that is the cone which is the envelope of the planes normal to each generator of the normal cone, touches the solution surface at each point. The solution surface is therefore the envelope of all these dual cones (which are also called *Monge cones*). For the quasilinear case the normal cone is a plane since $ap + bq = c$, and its dual degenerates into a line in the direction (a, b, c) which must be tangent to the solution surface. In the fully nonlinear case we may construct this Monge cone by finding the envelope of the planes $(x - x^*)p + (y - y^*)q = u - u^*$ through the point (x^*, y^*, u^*) and subject to $F(x^*, y^*, u^*, p, q) = 0$. As in §2.6, the envelope is given by

$$x - x^* = \lambda\frac{\partial F}{\partial p}, \quad y - y^* = \lambda\frac{\partial F}{\partial q},$$

where p, q and λ must be eliminated between these four relations. Eliminating λ, we obtain

$$\frac{x - x^*}{\partial F/\partial p} = \frac{y - y^*}{\partial F/\partial q} = \frac{u - u^*}{p\,\partial F/\partial p + q\,\partial F/\partial q},$$

and thus a small vector $(\delta x, \delta y, \delta u)$ lying in a characteristic strip satisfies

$$\frac{\delta x}{\partial F/\partial p} = \frac{\delta y}{\partial F/\partial q} = \frac{\delta u}{p\,\partial F/\partial p + q\,\partial F/\partial q} = \delta t,$$

say, for (x, y, u) near (x^*, y^*, u^*). We cannot eliminate p and q, but we can obtain expressions for δp and δq given by

$$\delta p = \frac{\partial p}{\partial x}\delta x + \frac{\partial p}{\partial y}\delta y = \delta t\left(-\frac{\partial F}{\partial x} - p\frac{\partial F}{\partial u}\right),$$
$$\delta q = \frac{\partial q}{\partial x}\delta x + \frac{\partial q}{\partial y}\delta y = \delta t\left(-\frac{\partial F}{\partial y} - q\frac{\partial F}{\partial u}\right),$$

using (8.7)–(8.11). Finally, letting $\delta t \to 0$, Charpit's equations (8.19) are recovered.

This discussion clearly reveals the existence of the so-called *integral conoids* formed by all the characteristics passing through a given point in (x, y, u) space. The integral conoid is the global extension of the Monge cone at that point and the solutions of the Cauchy problem for (8.6) can also be thought of geometrically as the envelope of all the integral conoids through the initial curve. As indicated in Fig. 8.1, this construction reveals the non-uniqueness we mentioned earlier in this section, and the situation could be far more complicated than that for the single-sheeted conoids that we have sketched.

A final general remark concerns those situations in which we may be unable to solve Charpit's equations yet we may be lucky enough to be able to guess a two-parameter family of solutions of (8.6) in the form $u = f(x, y; \alpha, \beta)$. Then it is easily verified that the eliminant of α between

$$u = f(x, y; \alpha, \beta(\alpha)) \quad \text{and} \quad \frac{\partial f}{\partial \alpha} + \frac{\mathrm{d}\beta}{\mathrm{d}\alpha}\frac{\partial f}{\partial \beta} = 0$$

is also a solution for any function $\beta(\alpha)$. Thus, if a solution to (8.6) with two arbitrary constants can be found, then many more solutions can be generated by different choices of $\beta(\alpha)$. Indeed, as mentioned in §1.9, we expect the general solution of any first-order partial differential equation with two independent variables to contain at least one arbitrary function of one variable in this way. In the quasilinear case considered in §1.5, the general solution was obtained by setting one constant of integration of the characteristic ordinary differential equation to be an

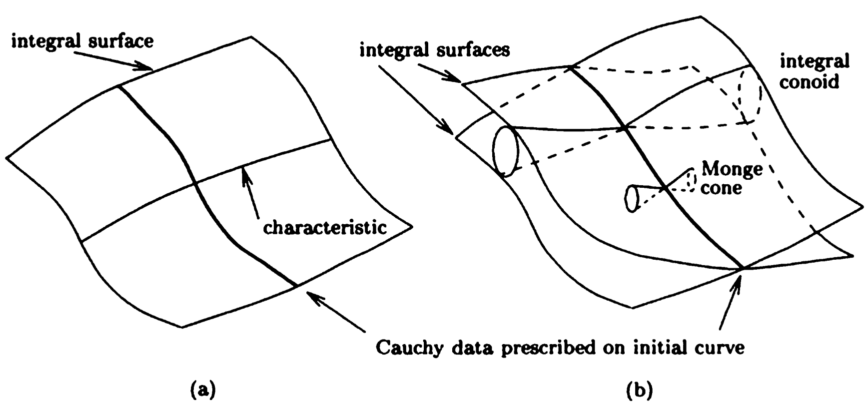


Fig. 8.1 Solution of the Cauchy problem: (a) the quasilinear case; (b) the non-quasilinear case.

arbitrary function of the other. In general, however, Charpit's equations demand that a more elaborate procedure be followed.

8.2.2 More independent variables

While the arguments above could never be extended to either scalar higher-order equations or vector equations, the generalisation to m independent variables is quite painless. Changing notation to

$$F(x_i, u, p_i) = 0, \qquad \text{where} \qquad p_i = \frac{\partial u}{\partial x_i} \quad \text{for } i = 1, \ldots, m, \tag{8.21}$$

Charpit's equations are simply

$$\dot{x}_i = \frac{\partial F}{\partial p_i}, \tag{8.22}$$

$$\dot{p}_i = -\frac{\partial F}{\partial x_i} - p_i \frac{\partial F}{\partial u}, \tag{8.23}$$

$$\dot{u} = \sum_1^m p_i \frac{\partial F}{\partial p_i}. \tag{8.24}$$

Anyone familiar with classical mechanics can now make an observation: if F is independent of u, then (8.22) and (8.23) are simply Hamilton's equations for a mechanical system in which x_i are generalised coordinates, p_i are generalised momenta and F is the Hamiltonian. Since it is easy to generalise our assertion that Charpit's equations give the solution of the partial differential equation to this case, we have thus derived the remarkable result that any problem in classical mechanics with a finite number of degrees of freedom is equivalent to a scalar first-order partial differential equation in which $\partial F/\partial u = 0$. We will explore the implications of this further in §8.3.

We remark that, even though the Cauchy problem involves data specified on an $(m-1)$-dimensional surface, the solution of (8.21) is still expressed in terms of ordinary differential equations which hold along the one-dimensional curves that we have called characteristics. This situation is a generalisation of that in §1.8 and hence, when $m > 2$, it is reasonable to call these one-dimensional curves *bicharacteristics*, reserving the adjective characteristic for surfaces of dimension $m - 1$.

The preceding theory paves the way for many exciting investigations into a great variety of problems in science and industry, and we now describe some of these, beginning, as always, with the simplest configurations.

8.2.3 The eikonal equation

Charpit's equations are greatly simplified when F does not depend explicitly on x, y and u. In this case, $\dot{p} = \dot{q} = 0$ on a characteristic, so that $p = p_0(s)$, $q = q_0(s)$ and the characteristic projections in the (x, y) plane, often called *rays*, are straight lines with slope $(\partial F/\partial q)_0 \,/\, (\partial F/\partial p)_0$ (as usual, the suffix zero denotes values on the initial curve). On these rays,

$$x = x_0 + t\left(\frac{\partial F}{\partial p}\right)_0, \quad y = y_0 + t\left(\frac{\partial F}{\partial q}\right)_0, \quad u = u_0 + t\left(p\frac{\partial F}{\partial p} + q\frac{\partial F}{\partial q}\right)_0, \quad (8.25)$$

and the solution may be obtained by first eliminating t and then s. The solution surface is a special case of a *ruled surface* since, at each point, there is a straight line lying in the surface; such surfaces will be discussed again in the final section of this chapter.

8.2.3.1 Sand piles

When (8.1) is used to model a sand pile on a horizontal base, the boundary conditions are $u = 0$ on $x = x_0(s)$, $y = y_0(s)$.[158] Thus $p_0x_0' + q_0y_0' = 0$ and $p_0^2 + q_0^2 = 1$, so that

$$p_0 = \frac{\pm y_0'}{((x_0')^2 + (y_0')^2)^{1/2}},$$

where the sign must be chosen so that the sand pile lies in $u > 0$. The ray equations (8.25) reduce to

$$x = 2tp_0 + x_0, \quad y = 2tq_0 + y_0, \quad u = 2t,$$

and, assuming the base has a smooth boundary, the solution can be obtained locally by eliminating s between $x - x_0 = up_0$ and $y - y_0 = uq_0$. This may not be at all simple to do explicitly and we consider two special cases.

For a circular base, we have

$$x_0 = \cos s, \quad y_0 = \sin s,$$

and we find

$$p_0 = -\cos s, \quad q_0 = -\sin s$$

and

$$u = 1 - (x^2 + y^2)^{1/2}.$$

Thus the sand pile is a cone of circular cross-section and apex $u = 1$, $x = y = 0$; it is in fact the integral conoid at this point, as shown in Fig. 8.2.

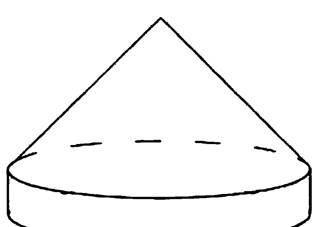


Fig. 8.2 Sand pile on a circular base.

[158]With these boundary conditions, the solution u is also the minimum distance from (x, y) to the curve $x = x_0(s)$, $y = y_0(s)$; see Exercise 8.2.

The solution breaks down at the apex, as may be confirmed by evaluating the Jacobian $|\partial(x,y)/\partial(s,t)|$ in the form

$$2py' - 2qx' = 2\big(2t(q_0'p_0 - q_0p_0') + y_0'p_0 - x_0'q_0\big) = 2(2t-1).$$

This vanishes when $t = 1/2$, $u = 1$ and $x = y = 0$, thus limiting the domain of definition to $t < 1/2$, the interior of the circle minus the centre.

For an elliptical base, we write $x_0 = a\cos s$ and $y_0 = b\sin s$, with $a > b$, and $p_0 = -\cos\theta$ and $q_0 = -\sin\theta$, where $b\tan\theta = a\tan s$. To avoid the algebraic complexity of a direct elimination of s and θ, we just consider the ray pattern. A typical ray has equation

$$y - b\sin s = \tan\theta(x - a\cos s), \quad a > b > 0, \tag{8.26}$$

and intersects $y = 0$ and, by symmetry, at least one other ray, where $x = \big((a^2 - b^2)/a\big)\cos s$. Thus the solution breaks down on $y = 0$, $|x| \leqslant (a^2 - b^2)/a$, and the surface has a 'ridge line'. The height of this ridge line is tedious to find, but at $x = 0$, where $s = \pi/2$ and $\theta = 3\pi/2$, $u = b$; at $x = (a^2 - b^2)/a$, $s = 0$ and $\theta = \pi$, so that $u = b^2/a$.

The discussion above has an intuitive component because of our bald assertion that a ridge line exists. This clearly accords with everyday experience with, say, a spoon heaped with fine dry sugar, but in the next example we will see that there are other mathematical solutions of (8.1) which do not 'stop' at the ridge. In fact, the ridges are reminiscent of the shocks we encountered in the quasilinear theory of Chapters 1 and 2, and we will have more to say about their mathematical status shortly.

8.2.3.2 Geometric optics

Since geometric optics is described by the same equation as that for sand piles, it might be thought that there is nothing more to say about it. However, the physical interpretation of the dependent variable is so different that a separate theoretical development is necessary.

The characteristic projections of (8.1) in the (x, y) plane, which are of course straight lines, are called ***light rays*** in the geometric optics applications. They are parallel to ∇u and hence are normal to the level sets of u which are the contour lines in the sand pile example. However, if we return to the time domain, as in (8.2), it is natural to define a ***wave-front*** as a level curve of the phase of ϕ.[159] Hence, remembering the definition of u in (8.4), if $u = u_0$ defines a wave-front at $\tau = \tau_0$, and (x_0, y_0) is a point on it, then, at a later time τ_1, the wave-front is given by $u = u_1 = u_0 + \tau_1 - \tau_0$; it is the projection in the (x, y) plane of the envelope of the integral conoids (Monge cones in this case) evaluated at $u = u_1$, namely

$$(x - x_0)^2 + (y - y_0)^2 = (u_1 - u_0)^2 = (\tau_1 - \tau_0)^2, \tag{8.27}$$

where x_0 and y_0 vary over the wave-front at $\tau = \tau_0$. As in Fig. 8.3, (8.27) represents a family of circles of prescribed radius and this is called ***Huygens' construction***. If

[159]This is in accord with the nomenclature in §4.6.

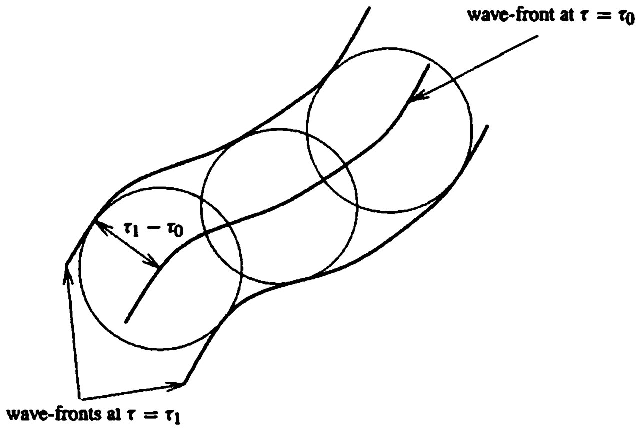


Fig. 8.3 Huygens' construction.

the wave speed is not constant, say if the medium has variable refractive index, then k is a function of position, the rays are curved and Huygens' construction only holds for small time intervals.

Two serious difficulties must be addressed before we can begin to use geometric optics to represent solutions of practical interest in wave propagation. The first concerns a phenomenon that we have already encountered in our study of sand piles, namely the intersection of characteristics. In optics, the interpretation of characteristic intersection differs fundamentally from that in Chapter 1 and at the end of the previous example. In both these latter situations, the dependent variable was required to be single-valued, and hence the intersection of characteristics motivated the introduction of a shock or ridge, respectively. However, for wave applications there is no restriction that the phase u should be single-valued nor even that it should be real (which leads to another topic to be mentioned later). Indeed, since there is no reason why many waves should not exist at any point, we can happily continue our characteristics through any intersection point, and in general they form an envelope[160] called a *caustic*. A famous example is the solution of Charpit's equations for $|\nabla u|^2 = 1$, with data

$$u_0 = s \quad \text{on} \quad x_0 = \cos s + \sin s, \ y_0 = \sin s - \cos s, \tag{8.28}$$

a circle of radius $\sqrt{2}$. We soon find that

$$\begin{aligned} p = p_0 &= -\sin s \quad \text{or} \quad p_0 = \cos s, \\ q = q_0 &= \cos s \quad \ \ \text{or} \quad q_0 = \sin s, \end{aligned}$$

and, redefining the parameter t so that $\mathrm{d}x/\mathrm{d}t = p$, etc.,

[160] In special cases, of course, the rays may 'focus' at a point, but such a geometry is not stable to small perturbations. Stable envelopes can be classified using catastrophe theory [1].

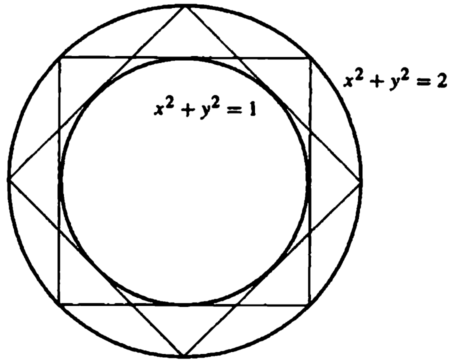


Fig. 8.4 Circular caustic in a circle.

$$x = \cos s - (t-1)\sin s \quad \text{or} \quad x = \sin s + (t+1)\cos s,$$
$$y = \sin s + (t-1)\cos s \quad \text{or} \quad y = -\cos s + (t+1)\sin s,$$
$$u = s + t,$$

respectively. There are two values of p_0 and q_0 corresponding to any value of s, and the rays envelop a circular caustic on $x^2 + y^2 = 1$, as shown in Fig. 8.4.

In this situation the caustic separates a region of greater illumination, $x^2+y^2 > 1$, through each point of which there are two rays, from one of less illumination, $x^2 + y^2 < 1$, where there are no real rays, and all caustics share this attribute of separating brighter regions from darker ones.[161] However, the one that is easiest to see in practice is the nephroid formed by sunlight reflecting from the curved vertical side of a nearly-filled coffee mug; here the bright region has four rays through any point and in the dark region there are just two rays through any point (see Exercise 8.8 and Fig. 8.8).

The bunching of the rays in Fig. 8.4 suggests that the region of greatest illumination is near the caustic. To see this more precisely, we note that, in general, the solution

$$x = p_0(s)t + x_0(s), \quad y = q_0(s)t + y_0(s), \quad u = t + u_0(s), \tag{8.29}$$

where

$$p_0^2 + q_0^2 = 1 \quad \text{and} \quad \frac{\mathrm{d}x_0}{\mathrm{d}s}p_0 + \frac{\mathrm{d}y_0}{\mathrm{d}s}q_0 = \frac{\mathrm{d}u_0}{\mathrm{d}s},$$

yields a Jacobian (see Exercise 8.11)

$$\left|\frac{\partial(x,y)}{\partial(s,t)}\right| = (q_0p_0' - p_0q_0')\,(t + T(s)),$$

where

[161]This is because, as we shall see, the field inside the caustic is exponentially small in the parameter k in (8.4).

$$T(s) = \frac{q_0 x_0' - p_0 y_0'}{q_0 p_0' - p_0 q_0'}.$$

Hence, unless $q_0 p_0' = p_0 q_0'$, which only happens for plane waves in which u is real and linear in x and y, there is always a caustic at $t = -T(s)$. Now this has serious implications for the amplitude A in (8.4). By equating terms of $O(k)$ to be zero in the ansatz there, A satisfies

$$A\nabla^2 u + 2\nabla u \cdot \nabla A = 0. \tag{8.30}$$

This linear partial differential equation for A has an explicit solution by the methods of Chapter 1; following Exercise 8.11, we find that

$$A(s) = A_0(s)\sqrt{\frac{T(s)}{t + T(s)}}, \tag{8.31}$$

indicating an increased illumination as $t \to -T(s)$. In the example (8.28), $T = -1$ and A has an inverse square-root singularity as $t \uparrow 1$. This is clearly a warning that the geometric optics ansatz has broken down and we must make a different high-frequency approximation to Helmholtz' equation near the caustic, but this would not be in the spirit of the book.[162]

A less symmetric example of a caustic has just been encountered in the elliptic sand pile solution (8.26). Continuing the characteristics through the ridge lines gives the ray picture shown in Fig. 8.5, and it soon becomes clear that caustics develop cusps whenever a boundary with constant Dirichlet data for u has a maximum or minimum in its curvature (see Exercise 8.9).

In summary, the problems created by intersecting characteristics for wave problems are more innocuous than when u is required to be single-valued. Although the caustics perturb the ray model locally, they do not affect the solution globally in the way that shocks or ridge lines do.

Our second serious difficulty with practical ray theory occurs when we consider configurations that involve physical boundaries. One of the commonest concerns the scattering of a plane wave, say with rays parallel to the x axis and wavefronts parallel to the y axis, by a smooth convex obstacle[163] (see Fig. 8.6) at which Dirichlet or Neumann conditions are imposed. This situation might arise in tomography, radar or ultrasonic testing and, at first sight, all we have to do is follow the familiar rules of optics and draw the reflected rays, as in Fig. 8.6.

[162]Those who may doubt the geometric optics approach may, in this case, look at the function

$$\psi = J_k\left(k\sqrt{x^2+y^2}\right)\exp\left(ik\tan^{-1}\left(\frac{y}{x}\right)\right),$$

which is an exact solution of Helmholtz' equation, and take the limit as $k \to \infty$; the solution corresponding to (8.28) is the asymptotic limit of $\psi e^{ik\pi/4}$ for $x^2 + y^2 > 1$, as long as we are not close to the caustic.

[163]When the scatterer is a circular cylinder, it is possible to write down the reflected field exactly by separating the variables. Unfortunately, the asymptotic form of the eigenfunction expansion, which involves $H_n^{(1)}(kr)\cos n\theta$, is not easy to find.

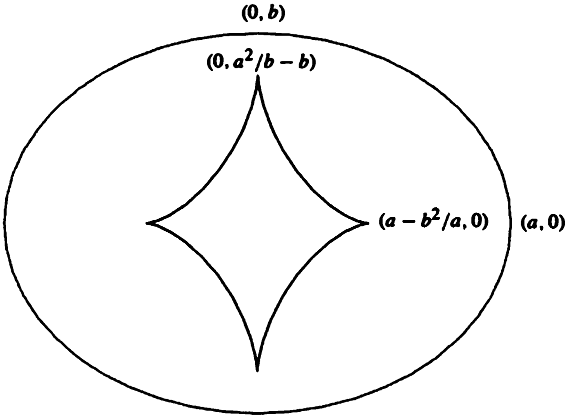


Fig. 8.5 Caustic in an ellipse.

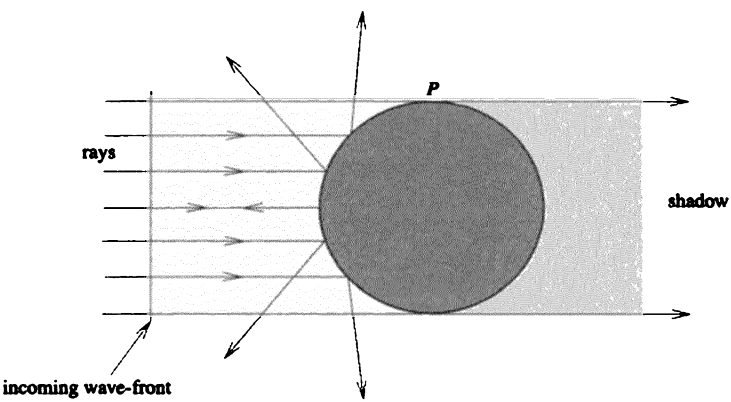


Fig. 8.6 Scattering of a plane wave by a circular cylinder.

The justification or otherwise of this procedure is a surprisingly complicated matter, requiring first the analysis of the basic reflection process. Suppose, say, we have Dirichlet data $\phi = 0$ and an incident wave e^{ikx}. When we write the total field as $e^{ikx} + \psi \sim e^{ikx} - Ae^{iu}$, at the boundary we have $A_0 e^{iku_0} = e^{ikx_0}$, i.e. $u_0 = x_0$ and $A_0 = 1$. The direction of the scattered ray is (p_0, q_0), which satisfy

$$\frac{du_0}{ds} = \frac{dx_0}{ds} = p_0 \frac{dx_0}{ds} + q_0 \frac{dy_0}{ds};$$

this says that the angle between the tangent to the scatterer and the reflected ray is the same as that between the tangent and the incident ray. This is called *specular reflection* and, although we have only derived this law for geometric optics, it happens to be obeyed by plane wave solutions of the full scalar wave equation.

This argument works in general unless the scatterer has corners or, as at P in Fig. 8.6, the ray is at 'grazing' incidence.[164] In either case, modifications have to be made to the geometric optics ansatz which are much too complicated to be discussed here, but form the basis of J. B. Keller's Geometric Theory of Diffraction [5]. The modifications are by no means local to the body, involving, for example, the shadow boundaries. Also, if the scatterer has any concavities, then caustics are almost inevitably present.

We conclude this brief discussion of a very large subject with three remarks. First, there is no reason why we should not apply a WKB-style ansatz to the full wave equation rather than considering the single Fourier component (8.2). There is then no large parameter k in which to expand asymptotically, but we can justify the procedure in cases where the wavelength is much smaller than the length scale over which we are considering the solution. Thus, writing

$$\psi \sim A e^{iu(x,y,\tau)},$$

where ∇u is large, we obtain

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 = \left(\frac{\partial u}{\partial \tau}\right)^2. \tag{8.32}$$

This equation can, of course, be solved by Charpit's method (see Exercise 8.12), but it is precisely the equation for the characteristic surfaces of the two-dimensional wave equation, as given in a different notation in §2.6. Thus, the geometric optics approach coincides with our theory of characteristics, as has already been evident from the discussion of rays and wave-fronts in this chapter and Chapter 4; the rays are simply the bicharacteristics associated with Charpit's method. However, geometric optics has the advantage of being susceptible to improvement by including the effects of longer wavelengths, because we can proceed to compute A in (8.4) as a series in inverse powers of k; the theory of characteristics is merely the 'zero-wavelength' limit.[165]

Secondly, there is no reason why we should not apply geometric optics to the modified Helmholtz equation

$$(\nabla^2 - k^2)\psi = 0. \tag{8.33}$$

We have already seen that this equation describes physical problems that are quite different from wave propagation, yet, by setting $\psi \sim Ae^{ku}$, we obtain (8.1) again! However, the difference between the physical interpretation of (8.33) and (8.3) has already emerged: for problems for which (8.33) is the correct model, u must be single-valued and hence the 'ridge line' solution would be the appropriate one for,

[164]Note that the breakdown of the ray solution at P can be likened to that near a caustic because two rays, one incident and one reflected, coincide at P.

[165]However, we must remember the situation mentioned in Chapter 4, where a highly localised initial condition for the wave equation could create an equally localised response near the characteristics emanating from the source, and also a non-trivial response away from these characteristics.

say, a chemical reaction occurring in an elliptical container, as in (8.26). Moreover, starting from (8.33), we could carry out a 'shock'-type analysis of the ridge structure in the same spirit as in Chapters 2, 6 and 7 (see Exercise 8.14).

Finally, we mention what is perhaps the most challenging aspect of geometric optics. This is the fact that u does not have to be real for (8.1) to apply to wave propagation, although a complex u would correspond to ψ being either exponentially large or exponentially small. We can see a hint of this in our simplest example (8.28) because, if s and t are such that $x^2 + y^2 < 1$, then it is easy to see that u is complex. But such values of s and t are also complex and this means that we are led to the idea of *complex rays*, in which x and y must also be complexified. Further discussion of this subject, which is not at all well developed, cannot be given here.

*8.2.4 Eigenvalue problems

The application of geometric optics to find high eigenfrequencies using Helmholtz or other wave-related equations turns out to be surprisingly difficult and delicate. We will content ourselves with describing one simple example which will be of great motivational value in the next section.

Suppose we wish to find large values of λ such that the problem

$$\nabla^2\psi + \lambda\psi = 0 \tag{8.34}$$

in the circle

$$x^2 + y^2 \leqslant R^2,$$

with either homogeneous Dirichlet or Neumann boundary data, has non-trivial solutions, i.e. $-\lambda$ is an eigenvalue of ∇^2. Let us assume that λ is large enough that we can approximate ψ in the form

$$\psi(x, y; \lambda) \sim A(x, y)e^{i\sqrt{\lambda}u(x,y)},$$

where u satisfies $|\nabla u|^2 = 1$. The phase u can only be determined from Charpit's equations if we know its value at the boundary. This data determines $\partial u/\partial\theta$ and $\pm\partial u/\partial r$ in polar coordinates. Since the ray is in the direction of ∇u, the components of ∇u are $(\sin\phi, \cos\phi)$, also in polar coordinates, where ϕ is the launch angle of the two possible rays through any point P on the boundary, as shown in Fig. 8.7. On each of these rays, near P, the two phases are

$$u_\pm \sim u_P + R(\theta - \theta_P)\cos\phi \pm (R - r)\sin\phi + \cdots.$$

We must satisfy the boundary condition $\psi = 0$, and, since we cannot do this by taking $A = 0$ on $r = R$ (as then A vanishes everywhere), we achieve this by setting

$$\psi(x, y) = A(x, y)\left(e^{i\sqrt{\lambda}u_+} - e^{i\sqrt{\lambda}u_-}\right).$$

The only potential problem with this procedure is that the '+' ray might, after a number of bounces as in Fig. 8.7(a), return to P as a '−' ray, bringing with it

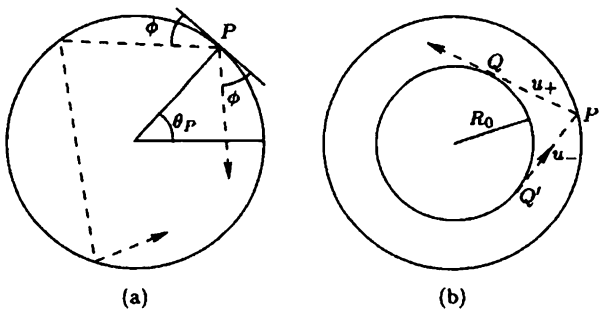


Fig. 8.7 Eigenvalues of the Laplacian in a circle.

a value of u_+, obtained by integrating Charpit's equations along this ray, that is inconsistent with u_- at P. This consistency requirement will lead to the determination of the eigenvalues. Now for most values of ϕ, namely those for which $\cos^{-1}(R_0/R)$ is an irrational multiple of π, where R_0 is as in Fig. 8.7(b), this does not occur. However, we must also recognise that the rays launched at the angle ϕ from different points on the boundary have an envelope (caustic) with radius R_0 (see Fig. 8.7(b)), which is itself a route along which the phase u_+ can travel. Thus we must consider trajectories such as that from P to Q, around the caustic any number of times ending at Q', and back to P. When we work out the change in u_+ along such a trajectory, it must be an integer multiple of 2π if we are to be consistent with u_- at P. After a simple calculation, this leads to the conditions

$$2\pi R_0\sqrt{\lambda} = 2\pi n_1, \tag{8.35}$$

$$2\sqrt{\lambda}\left(\sqrt{R^2 - R_0^2} - R_0\cos^{-1}\left(\frac{R_0}{R}\right)\right) = 2\pi n_2, \tag{8.36}$$

for some integers n_1 and n_2. Finally, we are led to the requirement that the eigenvalue λ is such that the eliminant of R_0 between equations (8.35) and (8.36) is satisfied for some integers n_1 and n_2. It can be shown that this gives a good approximation to the n_2th zero of the Bessel function $J_{n_1}(\sqrt{\lambda}R)$ that appears in the exact solution of (8.34).

Unfortunately, this is not an easy technique to use on more general problems. Even a slight modification to the boundary can lead to formidable geometric and analytical difficulties in ray tracing. Furthermore, any attempt to increase the accuracy of results by including the amplitude A as well as u becomes quite complicated. One inaccuracy in (8.35) is immediately apparent because we have assumed that u is continuous when a ray grazes the circle $r = R_0$. In fact, a local calculation of the type mentioned after (8.31) shows that u always changes by $\pi/2$ at a grazing point, and hence n_2 in (8.36) should at the very least be replaced by $n_2 + 1/4$.

One insight we may gain from this ray analysis concerns the qualitative behaviour of the eigenfunctions corresponding to the large eigenvalues. Figure 8.7(b) suggests that the eigenfunctions corresponding to the launch angle ϕ will be localised near the caustic $r = R_0$. Were we to try a similar analysis in a rectangular domain, the absence of any caustics suggests that the eigenfunctions are not localised in this geometry, as can be verified by considering their exact representation in terms of trigonometric functions.

8.2.5 Dispersion

The ideas described above are by no means confined to (8.1). Indeed, general models involving linear partial differential equations for wave propagation with constant coefficients have solutions in which the dependent variables are all proportional to $e^{i(\mathbf{k}\cdot\mathbf{x}-\omega\tau)}$, as long as some *dispersion relation*

$$F(\mathbf{k},\omega) = 0 \tag{8.37}$$

holds between the wavenumber vector $\mathbf{k}$ and the frequency ω. This approach to dispersion is simply a repetition of the ideas of §7.2.1; the fact that the model permits wave solutions at all means that there are real values of $\mathbf{k}$ and ω that satisfy (8.37). For linear problems, the property that a model admits a dispersion relation with real $\mathbf{k}$ and ω is another approach to hyperbolicity, as defined in §3.3. It has the advantage of applying to models, such as that for surface gravity waves in §7.2.1, which contain more information than just partial differential equations; for instance, F could be crucially dependent on the boundary conditions, as in the example below.

The theory of this chapter has a really useful role to play when we apply the WKB methodology to linear wave propagation models by writing all the dependent variables in the form Ae^{iku}. The highest derivatives are the only relevant terms because, as stated above, the restriction of the WKB approach is that the wavelength is much less than any length scale of interest, and thus it is a 'short wavelength' or 'far-field' approximation. In any case, the operator $\partial/\partial x_j$ is equivalent to multiplication by ik_j and (8.37) implies that u satisfies

$$F\left(\nabla u, -\frac{\partial u}{\partial \tau}\right) = 0. \tag{8.38}$$

For the simple example (8.32), it is easily seen from Charpit's equations that, not only are the components of ∇u and $\partial u/\partial\tau$ constant along characteristics, but also that u itself is constant. The fact that the phase u does not change along a characteristic (or ray) results from the fact that all disturbances, no matter what their frequency or wavelength, travel at the same speed. Hence there is no dispersion, i.e. no 'mixing' of waves of different wavelengths.

An example in which dispersion is important is the famous problem of the far-field of the wave pattern created by a ship travelling with speed V in the x direction on an infinitely deep ocean. As shown in Exercise 8.15, the dispersion relation is

$$(-\omega + Vk_1)^4 = g^2(k_1^2 + k_2^2), \tag{8.39}$$

where x and y are horizontal coordinates in the ocean surface, g is the acceleration due to gravity and $\mathbf{k} = (k_1, k_2)$. Hence, in the steady state in which $\omega = 0$, the WKB approximation for the phase u of the surface elevation satisfies

$$V^4p^4 = g^2(p^2 + q^2), \tag{8.40}$$

where, as usual, $p = \partial u/\partial x$ and $q = \partial u/\partial y$. Now, in the WKB approximation, the ship is effectively a point, say the origin, and hence we are only interested in the integral conoid through it, which is given by

$$x = (4V^4p^3 - 2g^2p)t, \quad y = -2g^2qt, \quad u = 2V^4p^4t,$$

together with (8.40). The most striking feature emerges when we consider the ray slope, which is $-pq/(p^2 + 2q^2)$ for *any* value of V (or g), and has a maximum of $1/(2\sqrt{2})$. Hence the waves are confined to a wedge of semi-angle $\sin^{-1}(1/3)$, within which there are two curved wave-fronts through any point touching on the wedge, each carrying a different phase. The limiting straight rays are caustics at which the Jacobian vanishes, and there a large-magnitude 'bow wave' is seen, as shown in the photograph in [45, p. 117].

8.2.6 Bicharacteristics

One bonus of the m-dimensional generalisation of the situation of Fig. 8.1 is that we are now in a position to construct the bicharacteristics of m-dimensional hyperbolic equations, as introduced in §2.6, of the form

$$\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \mathbf{u}}{\partial x_i} = \mathbf{c}.$$

Indeed, the fact that the characteristics satisfying (8.22)–(8.24) yield the generators of the Monge cone locally means that we only have to find the characteristics of (2.53), namely

$$Q\left(\frac{\partial \phi}{\partial x_1}, \dots, \frac{\partial \phi}{\partial x_m}\right) = \det\left(\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \phi}{\partial x_i}\right) = 0.$$

Thus, for the wave equation

$$\frac{\partial^2 u}{\partial t^2} = a_0^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right),$$

we need the characteristics of

$$a_0^2\left(\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial y}\right)^2\right) = \left(\frac{\partial \phi}{\partial t}\right)^2,$$

which, with $a_0 = 1$, is (8.32). For the characteristic cone through the origin, these are simply the generators

$$x = p\tau, \quad y = q\tau, \quad t = r\tau, \quad \text{where} \quad p^2 + q^2 = a_0^2 r^2,$$

but the geometry can easily become more complicated (see Exercise 8.13).

*8.3 Hamilton–Jacobi equations and quantum mechanics

The purpose of this section is to give a brief account of how the theory of partial differential equations holds the key to the link between classical and quantum mechanics. It is of course an intricate story and we will only give the very simplest outline of what is involved.

We begin by recalling that many classical mechanics problems with a finite number of degrees of freedom can be recast as Hamilton's equations

$$\dot{p}_i = \frac{-\partial H(q_i, p_i)}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H(q_i, p_i)}{\partial p_i} \quad \text{for } i = 1, \ldots, n,$$

for the generalised coordinates $q_i(t)$ and generalised momenta $p_i(t)$. In §8.2.2 we have already remarked that these equations are Charpit's equations for the following differential equation for u:

$$H(\mathbf{q}, \mathbf{p}) = 0, \quad \mathbf{q} = \{q_i\}, \quad \mathbf{p} = \{p_i\} = \left\{ \frac{\partial u}{\partial q_i} \right\}; \tag{8.41}$$

we simply have to replace $\mathbf{x}$ in (8.21) by $\mathbf{q}$ to conform with popular convention and to suppress the dependence of $\mathbf{p}$ and $\mathbf{q}$ on time t. Hence we have a route from Newton's laws via Charpit's equation to (8.41), which is called the *Hamilton–Jacobi equation*. However, both Newton's laws and the Hamilton–Jacobi equation are equivalent to the *principle of least action*, namely that the *action*

$$u = \int_0^T L(\mathbf{q}, \dot{\mathbf{q}})\, \mathrm{d}t \tag{8.42}$$

is minimised over all possible $\mathbf{q}$ when the integration is taken along a solution of Newton's equations (i.e. along a ray of Charpit's equations), assuming $\mathbf{q}(0)$ and $\mathbf{q}(T)$ are prescribed. Here L is called the *Lagrangian* and the equivalence of (8.42) and Newton's laws is given in [46]. It is interesting to remark that the equivalence of (8.42) to the Hamilton–Jacobi equation (8.41) is analogous to the equivalence between Fermat's principle[166] and the eikonal equation (8.1) (see Exercise 8.17).

We now ask ourselves 'what partial differential equations may give rise to the Hamilton–Jacobi equation in the WKB limit, in the way that Helmholtz' equation gave rise to (8.1)?' This question clearly has no unique answer, since the WKB approximation does not usually 'balance' the highest derivatives and all the lower-derivative terms in any given linear differential equation, but one possibility is the Schrödinger equation

$$H\left(q_i, -\mathrm{i}\hbar \frac{\partial}{\partial q_i}\right) \psi(q_i) = E\psi(q_i). \tag{8.43}$$

Here ψ is called the wave function and E is a constant, the energy level of the system; $\hbar$ is also a constant which is very small when we work on scales much

[166]This states that the travel time (which is equivalent to arc length in a homogeneous medium) adopted by a light ray is a minimum compared to other geometrically possible paths.

greater than the size of an atom (i.e. much greater than 10 ångströms or so). We require the dependence of $H(\mathbf{q},\mathbf{p})$ on its last n arguments to be a linear combination of powers, giving for example

$$H\left(q_i, -\mathrm{i}\hbar\frac{\partial}{\partial q_i}\right) = -\frac{\hbar^2}{2m}\sum_{i=1}^{n}\frac{\partial^2}{\partial q_i^2} + V(q_i), \tag{8.44}$$

which is the Hamiltonian of a single particle of mass m moving in a potential V in n dimensions. The WKB ansatz $\psi \sim A\mathrm{e}^{\mathrm{i}u(q_i)/\hbar}$ can be expected to yield a good approximation on a scale of many ångströms and it gives

$$H(q_i, p_i) = E,$$

which, since E is constant, is equivalent to (8.41) as long as we identify the dependent variable with the phase u of the wave function ψ.[167] Thus the relation of the wave function in quantum mechanics to the trajectories of classical mechanics is analogous to that between the electromagnetic field and the light rays of geometric optics; both hinge on the role played by the phase of the wave function. Note that, when we identify Hamilton's equations of mechanics with Charpit's equations for the Hamilton–Jacobi equation (8.41), the parameter t along the characteristics must be taken to be real time. Indeed, T in (8.42) is real time in the Lagrangian formulation. However, the parameter along the characteristics of the eikonal equation (8.1) cannot be thus identified because (8.1) only applies to the spatial description of a single high-frequency component in the solution of the wave equation. Indeed, it is helpful to think of the ray parameter in this case as a weighted distance, as explained in Exercise 8.17.

A quantity of primary practical interest in quantum mechanics is the spectrum of energy levels E, which emerge as eigenvalues in this theory. We can attempt to find the energy levels for small $\hbar$ by generalising the argument of §8.2.4, which corresponds to a very simple case in which $V = 0$ in two dimensions. To do this we must generalise the all-important conditions that the change of phase along any closed path, in particular a ray path, if one can be found, is an integer multiple of $2\pi n\hbar$. Now we already know u in principle, assuming we have been able to integrate Hamilton's equations because, by Charpit's method,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \sum_{i=1}^{n} p_i\frac{\partial H}{\partial p_i} = \sum_{i=1}^{n} p_i\frac{\mathrm{d}q_i}{\mathrm{d}t},$$

so

$$u = u_0 + \int_0^t \sum_{i=1}^{n} p_i\,\mathrm{d}q_i, \tag{8.45}$$

where u_0 is the value of the phase on the boundary. It is a well-known result of classical mechanics that the minimisation of the action (8.42) yields (8.45),

[167] Unfortunately, when there are many particles, each has to have its own wave function ψ and the dimensionality of the resulting differential equation is the number of degrees of freedom of the system, which may be enormous.

and hence we can also identify the vitally important phase function u with the minimising action. Moreover, in principle we only have to ensure that

$$\oint \sum_{i=1}^{n} p_i \, \mathrm{d}q_i,$$

which is the change of phase along any closed path, is $2n\pi\hbar$. As in §8.2.4, even when we are fortunate enough to be able to take this integral along a closed ray path, care must be taken when evaluating the contributions from regions where the path touches an envelope. Unfortunately, this may be a laborious task, except in very simple geometries, because there will be many more 'families' u_i than the two we encountered in geometric optics.

*8.4 Higher-order equations

A remarkable number of phenomena have been illuminated by using non-quasilinear equations directly, and many of them would have been obscured had we prosaically differentiated the equations into quasilinear systems. Indeed, many aspects of the topics that we have mentioned richly deserve to be taken further, but most of them would require a considerable background knowledge of asymptotics.

Another reason for not proceeding further with non-quasilinear equations is that none of the results mentioned so far can be generalised to equations other than scalar first-order ones. Such generalised equations are relatively rare, but one interesting example is the *Monge–Ampère equation* of differential geometry, defined below. In its simplest form it concerns developable surfaces, i.e. ones that are envelopes of a one-parameter family of planes. Such surfaces are of great practical importance, say, in making curved sheets of glass or metal by bending plane sheets, to which they are isometric. They are special cases of ruled surfaces, as mentioned after (8.25). A simple paper bending experiment shows that a developable surface is necessarily the envelope of a one-parameter family of planes, say

$$p(\lambda)x + q(\lambda)y = u + \lambda \tag{8.46}$$

as λ varies. Hence it is the eliminant of λ between (8.46) and

$$p'(\lambda)x + q'(\lambda)y = 1.$$

Now (8.46) is a Legendre or 'contact' transformation, as in §4.8, and

$$\frac{\partial u}{\partial x} = p + \left(xp'(\lambda) + yq'(\lambda) - 1\right)\frac{\partial \lambda}{\partial x} = p,$$

and similarly $\partial u/\partial y = q$. Thus, since p and q are both functions of one variable λ,

$$\frac{\partial u}{\partial x} = f\left(\frac{\partial u}{\partial y}\right) \tag{8.47}$$

for some function f, and we have an equation that can be solved by (8.25). It can also be observed that (8.47) is the general first integral of the simplest form of the Monge–Ampère equation

$$rt - s^2 = 0, \tag{8.48}$$

where $r = \partial^2 u/\partial x^2$, $s = \partial^2 u/\partial x \partial y$ and $t = \partial^2 u/\partial y^2$. This equation has the geometric interpretation that the Gaussian curvature, which is the square root of the product of the two principal curvatures of the surface $u = u(x, y)$, is zero.

Now, in the spirit of this book, if we were confronted directly with (8.48), the best we could do would be to convert it into a quasilinear 3×3 system such as

$$r\frac{\partial t}{\partial y} + t\frac{\partial r}{\partial y} - 2s\frac{\partial s}{\partial y} = 0,$$
$$t\frac{\partial r}{\partial x} + r\frac{\partial s}{\partial y} - 2s\frac{\partial r}{\partial y} = 0,$$
$$\frac{\partial r}{\partial y} - \frac{\partial s}{\partial x} = 0,$$

which can easily be shown to possess just one double characteristic on which

$$t\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + 2s\frac{\mathrm{d}y}{\mathrm{d}x} + r = 0, \quad \text{i.e.} \quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{s}{t} = -\frac{r}{s}, \tag{8.49}$$

and also on which the Riemann invariant s/t (or r/s) is constant. Thus the characteristic is always straight. Moreover, we can interpret it geometrically by considering the second fundamental form

$$r\,\mathrm{d}x^2 + 2s\,\mathrm{d}x\,\mathrm{d}y + t\,\mathrm{d}y^2 \tag{8.50}$$

for the surface $u = u(x, y)$. This form determines the curvature of the normal cross-section of the surface in the direction $(\mathrm{d}x, \mathrm{d}y)$ and, for a developable surface, this curvature is zero on the (straight) generator through any point. Hence (8.49) implies that the characteristic through any point in the (x, y) plane is simply the projection of this generator.

Our final geometrical remark concerns the differential equation satisfied by a general ruled surface, not just a developable one. The condition for the surface $u = u(x, y)$ to contain a straight line at every point is that, for any (x, y), there are constants λ and μ such that

$$u(x + \tau, y + \lambda\tau) \equiv u(x, y) + \mu\tau$$

for all τ. Differentiation with respect to τ reveals that

$$\frac{\partial u}{\partial x} + \lambda\frac{\partial u}{\partial y} = \mu,$$

the arguments of the derivatives still being $(x + \tau, y + \lambda\tau)$, and we soon find that

$$\frac{\partial^2 u}{\partial x^2} + 2\lambda\frac{\partial^2 u}{\partial x \partial y} + \lambda^2\frac{\partial^2 u}{\partial y^2} = 0,$$

so that $\lambda = \left(-s \pm \sqrt{s^2 - rt}\right)/t$, the arguments again being taken at $(x+\tau, y+\lambda\tau)$. A final differentiation with respect to τ gives that u satisfies the third-order quasilinear equation

$$\frac{\partial \lambda}{\partial x} + \lambda \frac{\partial \lambda}{\partial y} = 0,$$

and it is a simple exercise to show that this equation is always satisfied by any solution of (8.48).

We conclude this section with an interesting postscript to our discussion of gas dynamics in §4.8. In the special case of inviscid *axisymmetric* flow, the relevant generalisation of (2.5)–(2.7) can be shown to lead not to (4.85) but rather to

$$F\frac{\partial^2 u}{\partial x^2} + G\frac{\partial^2 u}{\partial x \partial y} + H\frac{\partial^2 u}{\partial y^2} + \frac{1}{y}\frac{\partial u}{\partial y} = 0,$$

where again F, G and H are functions only of $\partial u/\partial x = p$ and $\partial u/\partial y = q$ (y is now a cylindrical polar coordinate). When we now use the Legendre transformation

$$w(p,q) = x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} - u,$$

so that $x = \partial w/\partial p$ and $y = \partial w/\partial q$, we obtain

$$F\frac{\partial^2 w}{\partial q^2} - G\frac{\partial^2 w}{\partial p \partial q} + H\frac{\partial^2 w}{\partial p^2} + \frac{q}{\partial w/\partial q}\left(\frac{\partial^2 w}{\partial p^2}\frac{\partial^2 w}{\partial q^2} - \left(\frac{\partial^2 w}{\partial p \partial q}\right)^2\right) = 0.$$

Hence we are led to an 'inhomogeneous' Monge–Ampère equation in which $rt - s^2$ (where now $r = \partial^2 w/\partial q^2$ and so on) is non-zero. We observe that, if $rt - s^2$ is a *known* function of position, so that the right-hand side of (8.48) is a prescribed function of x and y, then the calculation leading to (8.49) is the same, and, moreover, when the discriminant of the quadratic for the characteristics, which is equal to this function, is negative there are no real characteristics, while when it is positive there are two. The adjectives elliptic and hyperbolic are especially appropriate here because the former case is now associated with a surface of positive Gaussian curvature (for example, a sphere) and the latter with one of negative Gaussian curvature (for example, a saddle). For positive Gaussian curvature, the fundamental form (8.50) never vanishes. For negative Gaussian curvature, it vanishes in two directions $(\mathrm{d}x, \mathrm{d}y)$ called the *asymptotic lines*, which are lines that are bisected by the directions of maximum and minimum curvature. The case $rt - s^2 = 0$ can be described as parabolic; the double characteristics are the asymptotic lines. This agrees with our experience of paper bending; if, say, we constrain one end of a rectangular sheet to lie on a circle, then the paper forms a cylinder whose generators are the asymptotic lines. Likewise, a sector of a circular annulus can be bent into a cone and, indeed, inspection of a lightly-crumpled sheet of paper suggests that its displacement is made up of a large number of roughly conical patches joined along creases which we may interpret as 'ridge lines' for the solution.

More interestingly, equations of the type

$$rt - s^2 = Ar + Bs + Ct, \tag{8.51}$$

where A, B and C depend only on the first derivatives of the dependent variable, have been discussed in detail in [12]. They have the remarkable property that (8.49) becomes

$$(t - A)\left(\frac{dy}{dx}\right)^2 + (2s + B)\frac{dy}{dx} + r - C = 0,$$

and it is easy to see that there are two or no real characteristics according as

$$B^2 > 4AC \quad \text{or} \quad B^2 < 4AC,$$

respectively, the type of (8.51) is again determined only by its right-hand side.

All these phenomena are related to the fact that $rt - s^2 = 0$ is the Euler–Lagrange equation for, say, $\iint (p^2t + 2pqs + q^2r)\,dx\,dy$. One can always hope for there to be some special structure for any equation in which the non-quasilinear terms are invariants of the *Hessian matrix* $(\partial^2 u/\partial x_i \partial x_j)$; after all, the Laplacian is just the trace of this matrix.

Exercises

[Take $p = \partial u/\partial x$ and $q = \partial u/\partial y$ throughout.]

8.1. When sand is piled as high as possible on a table, its surface is $u = u(x, y)$, where $p^2 + q^2 = 1$. The table is a tilted rectangle so that

$$u(x,0) = 0 \quad \text{for } 0 < x < 1, \qquad u(x, a\cos\alpha) = a\sin\alpha \quad \text{for } 0 < x < 1,$$
$$u(0,y) = u(1,y) = y\tan\alpha \quad \text{for } 0 < y < a\cos\alpha,$$

where $0 < \alpha < \pi/4$ and $a \geqslant 1$. Solve Charpit's equations in a triangular region adjacent to each edge of the table to show that the surface consists of the planes

$$u = y, \quad u = y\tan\alpha + x\sqrt{1 - \tan^2\alpha},$$
$$u = y\tan\alpha + (1 - x)\sqrt{1 - \tan^2\alpha}, \quad u = a(\cos\alpha + \sin\alpha) - y,$$

which intersect at ridge lines. Show that the highest point of the pile is $u = a(\cos\alpha + \sin\alpha)/2$ as long as $a\cos\alpha \leqslant 1/\sqrt{1 - \tan^2\alpha}$.
What would happen if the table were L-shaped and horizontal?

8.2. Let $d(x, y)$ be the shortest distance from the point (x, y) to the smooth curve $y = f(x)$. Show that

$$d^2 = (x - X)^2 + (y - f(X))^2,$$

where

$$x - X + (y - f(X))\,f'(X) = 0.$$

Deduce that these formulæ provide the general solution of

$$\left(\frac{\partial d}{\partial x}\right)^2 + \left(\frac{\partial d}{\partial y}\right)^2 = 1.$$

You can see this without differentiation by taking the origin at (x, y) with the x axis along the normal from (x, y) to the curve, showing that $\partial d/\partial y = 0$, and noting that $|\nabla d|^2$ is invariant under rotation of the axes.

8.3. Suppose that u satisfies *Clairaut's equation*

$$u = xp + yq - f(p, q).$$

Show that, if $u = u_0(s)$ on $x = x_0(s)$, $y = y_0(s)$, then

$$x = x_0 e^t + \frac{\partial f}{\partial p}(p_0, q_0)(e^t - 1),$$
$$y = y_0 e^t + \frac{\partial f}{\partial q}(p_0, q_0)(e^t - 1),$$
$$u = u_0 e^t - \left(p_0 \frac{\partial f}{\partial p}(p_0, q_0) + q_0 \frac{\partial f}{\partial q}(p_0, q_0) - f(p_0, q_0)\right)(e^t - 1),$$

where $p_0(s)$ and $q_0(s)$ satisfy

$$u_0' = p_0 x_0' + q_0 y_0', \quad u_0 = x_0 p_0 + y_0 q_0 - f(p_0, q_0).$$

Noting that, when p and q are constant, Clairaut's equation is that of a plane in (u, x, y) space, deduce that $u = \alpha x + \beta y - f(\alpha, \beta)$ is a solution for any constants α and β. Hence show that the general solution is the result of eliminating α between

$$u = \alpha x + F(\alpha) y - f(\alpha, F(\alpha)),$$
$$x + \frac{dF}{d\alpha}\left(y - \frac{\partial f}{\partial q}(\alpha, F(\alpha))\right) - \frac{\partial f}{\partial p}(\alpha, F(\alpha)) = 0,$$

where F is arbitrary.

Remark. This is the only first-order equation whose general solution can be written down without any integrations, by virtue of the Legendre transformation (4.86).

8.4. Suppose that $pq = 1$ with $u = 0$ on $x + y = 1$. Use Charpit's method to show that $u = x + y - 1$ or $u = -(x + y - 1)$. Show also that, if $u = u_0(s)$ on $x = x_0(s)$, $y = y_0(s)$, then a solution only exists if $u_0'^2 > 4x_0' y_0'$.
By finding the solution for which all the characteristics pass through $x = y = 0$, show that the integral conoid through the origin is $u^2 = 4xy$.

8.5. Suppose that $u = f(x, y, p, q)$. Show that p and q satisfy the 2×2 system

$$\frac{\partial f}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial f}{\partial q}\frac{\partial q}{\partial x} = p - \frac{\partial f}{\partial x},$$
$$\frac{\partial f}{\partial p}\frac{\partial p}{\partial y} + \frac{\partial f}{\partial q}\frac{\partial q}{\partial y} = q - \frac{\partial f}{\partial y}.$$

Deduce that the characteristics are given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\partial f}{\partial q}\Big/\frac{\partial f}{\partial p} \quad \text{twice,}$$

and use the Fredholm Alternative to show that, along a characteristic,

$$\frac{\mathrm{d}x}{\partial f/\partial p} = \frac{\mathrm{d}y}{\partial f/\partial q} = \frac{\mathrm{d}p}{p - \partial f/\partial x} = \frac{\mathrm{d}q}{q - \partial f/\partial y}.$$

Deduce from the differential equation that these ratios are also equal to

$$\mathrm{d}u\Big/\left(p\frac{\partial f}{\partial p} + q\frac{\partial f}{\partial q}\right).$$

*8.6. Suppose that

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2} \quad \text{for } 0 < x < 1,$$

with

$$u(x,0) = 0, \quad u(0,\tau) = 1, \quad u(1,\tau) = 100.$$

Use a WKB ansatz $u \sim a(x,\tau)\mathrm{e}^{-v(x)/\tau}$, or the similarity solution (6.45), to indicate that, despite the different wall temperatures, the location at $\tau = 0+$ of the minimum value of u is at $x = 1/2$.

8.7. Check that $u = -x$ satisfies the equation of geometric optics and that it represents a plane wave of light incident from $x = +\infty$. When this light is incident on a parabolic reflector $y^2 = 4x$, the reflected field has to satisfy $u = -x$ on the parabola. Denote the values of p and q on the parabola by $p_0 = \cos s$ and $q_0 = \sin s$, and show that $u = u_0 = -x_0(s)$, $x = x_0(s)$, $y = y_0(s)$ satisfy

$$-x_0' = p_0 x_0' + q_0 y_0', \quad y_0 y_0' = 2x_0'.$$

Deduce that $y_0 = -2\tan(s/2)$ and $x_0 = \tan^2(s/2)$, and hence that the reflected rays are

$$y = (x-1)\tan s,$$

so that they all pass through the focus $(0, 1)$.

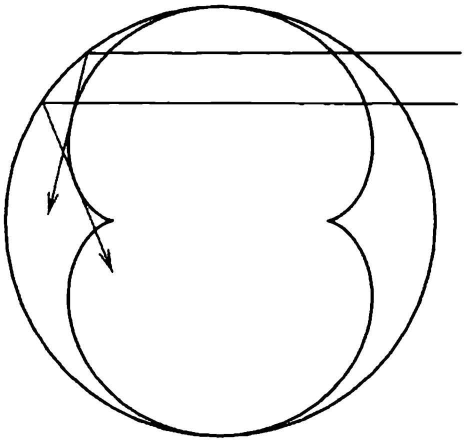


Fig. 8.8 The nephroid caustic in a circle; only the left-hand half is illuminated by the rays from $x = +\infty$.

*8.8. Suppose the reflector in Exercise 8.7 is $x^2 + y^2 = 1$, $x < 0$. Show that the Cauchy data is

$$u = -x_0(s), \quad x = x_0(s) = \cos\frac{s}{2}, \quad y = y_0(s) = \sin\frac{s}{2} \quad \text{for } \pi < s < 3\pi.$$

Hence show that the reflected rays are

$$x\sin s - y\cos s = \sin\frac{s}{2}.$$

Show that, for $s - 2\pi = \epsilon$, $|\epsilon| \ll 1$, the reflected rays all pass close to $x = -\frac{1}{2}$, $y = 0$, and that, when terms of $O(\epsilon^4)$ are neglected,

$$\left(x + \frac{1}{2}\right)\epsilon + \frac{\epsilon^3}{16} - y = 0.$$

Deduce that near $(-\frac{1}{2}, 0)$ the envelope of the rays is, approximately,

$$y = \pm\frac{8}{3^{3/2}}\left(-(x + 1/2)\right)^{3/2}.$$

The full envelope, which is called a *nephroid* from its resemblance to a kidney, is shown in Fig. 8.8.

8.9. Show that, when the eikonal equation is solved inside the ellipse $x^2/a^2 + y^2/b^2 = 1$, with $u = 0$ on the ellipse, the rays are normal to the boundary and their envelope is the caustic shown in Fig. 8.5, namely

$$x = \frac{a^2 - b^2}{a}\cos^3 s, \quad y = \frac{a^2 - b^2}{b}\sin^3 s \quad \text{for } 0 \leqslant s < 2\pi.$$

8.10. Suppose that the temperature $T(\mathbf{x})$ satisfies the convection–diffusion equation

$$(\mathbf{v}\cdot\nabla)T = \epsilon\nabla^2 T,$$

where $\mathbf{v}(\mathbf{x})$ is a prescribed velocity field. Show that, if we seek a WKB solution in which $T \sim Ae^{u(\mathbf{x})/\epsilon}$, then

$$(\mathbf{v}\cdot\nabla)u = |\nabla u|^2.$$

Show also that Charpit's equations are

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v} - 2\mathbf{p},\quad \frac{\mathrm{d}u}{\mathrm{d}t} = -\mathbf{p}\cdot\mathbf{v},\quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -(\mathbf{p}\cdot\nabla)\mathbf{v} + (\nabla\wedge\mathbf{v})\wedge\mathbf{p},$$

where $\mathbf{p} = \nabla u$, and deduce that to lowest order the isotherms T = constant bisect the angle between the characteristics and the streamlines, which are parallel to $\mathbf{v}$.

8.11. (i) Let t be a variable along a ray for the eikonal equation $p^2 + q^2 = 1$, scaled so that $\mathrm{d}x/\mathrm{d}t = p$, etc., and so

$$x = p_0 t + x_0,\quad y = q_0 t + y_0.$$

Define

$$J = \frac{\partial(x,y)}{\partial(s,t)} = \begin{pmatrix} x_0' + tp_0' & y_0' + tq_0' \\ p_0 & q_0 \end{pmatrix}.$$

Show that

$$|J| = (q_0 p_0' - p_0 q_0')\,(t + T(s)),$$

where, as on p. 371,

$$T(s) = \frac{q_0 x_0' - p_0 y_0'}{q_0 p_0' - p_0 q_0'}.$$

Show from (8.30) that

$$\frac{2}{A}\frac{\partial A}{\partial t} = -\nabla^2 u.$$

Invert J and use the relation

$$\begin{aligned}\nabla^2 u &= \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y}\\ &= p_0'\frac{\partial s}{\partial x} + q_0'\frac{\partial s}{\partial y}\end{aligned}$$

to show that $\nabla^2 u = J^{-1}\partial J/\partial t$. Deduce that

$$\frac{\partial}{\partial t}\left(A^2 J\right) = 0,\quad \nabla^2 u = \frac{1}{T+t},$$

where T is defined after (8.29). Finally deduce (8.31).

(ii) Suppose that such a ray extends to infinity. Show that

$$A^2 \sim \frac{A_0^2 T}{(x^2+y^2)^{1/2}},$$

so that the 'directivity' of the field radiated to infinity is $A_0^2 T$. (We used this result in §5.6.2.)

(iii) Suppose that $u = 0$ and $A = 1$ on the initial curve Γ. Show that on each ray

$$A = \sqrt{\frac{\rho(s)}{\rho(s)+t}},$$

where $\rho(s)$ is the radius of curvature of Γ at the point at which the ray leaves it. (It will help to take s to be arc length along Γ, so that the curvature is $1/\rho = y_0''x_0' - x_0''y_0'$.) Show that this implies that, in the absence of caustics, as $t \to \infty$,

$$A \to \frac{A_0}{r^{1/2}}\left(\frac{q_0x_0' - p_0y_0'}{q_0p_0' - p_0q_0'}\right)^{1/2},$$

where $A_0(s)$ is the initial amplitude. Deduce that, when $u_0(s) = 1$, $Ar^{1/2}$ is eventually inversely proportional to the curvature at the launch point. (In three dimensions it can be shown that it is the inverse of the Gaussian curvature that controls the far-field directivity; hence it is necessary to solve a Monge–Ampère equation in order to deduce the shape of the body from its scattered field.)

8.12. Suppose that a_0 is constant and

$$a_0^2\left(\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right) = \left(\frac{\partial u}{\partial \tau}\right)^2,$$

with Cauchy data $u = u_0(x,y)$ at $\tau = 0$. Show that p_0, q_0 and v_0 (where $v = \partial u/\partial \tau$) are given by

$$p_0 = \frac{\partial u_0}{\partial x}, \quad q_0 = \frac{\partial u_0}{\partial y}, \quad v_0 = \frac{\pm\sqrt{p_0^2+q_0^2}}{a_0}.$$

Deduce that the solution is

$$u(x,y,\tau) = u_0(X,Y),$$

where

$$x = 2p_0t + X, \quad y = 2q_0t + Y, \quad \tau = -\frac{2v_0t}{a_0^2}.$$

If u_0 is localised near $x = y = 0$, show that the solution is localised near the circle $x^2 + y^2 = a_0^2t^2$.

Remark. When used as an approximation to the solution of

$$\frac{\partial^2 \psi}{\partial t^2} = a_0^2 \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \right),$$

this result appears to violate Huygens' principle (see §4.6.2). However, the WKB approximation only describes very rapidly varying solutions and does not pick up more gradual variations inside the characteristic cone $x^2 + y^2 = a_0^2 t^2$.

*8.13. For the 'inhomogeneous wave equation'

$$\frac{\partial^2 \phi}{\partial t^2} = a_0^2(y) \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right),$$

which is a model for sound propagation in an ocean or atmosphere whose density varies vertically, show that the ray cones are $\phi(x, y, t) =$ constant, where

$$a_0^2(y) \left(\left(\frac{\partial \phi}{\partial x} \right)^2 + \left(\frac{\partial \phi}{\partial y} \right)^2 \right) = \left(\frac{\partial \phi}{\partial t} \right)^2,$$

and that the bicharacteristics satisfy

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = p, \quad \frac{\mathrm{d}y}{\mathrm{d}\tau} = q, \qquad \frac{\mathrm{d}t}{\mathrm{d}\tau} = -\frac{r}{a_0^2},$$

$$\frac{\mathrm{d}p}{\mathrm{d}\tau} = 0, \quad \frac{\mathrm{d}q}{\mathrm{d}\tau} = \frac{r^2}{2} \frac{\mathrm{d}}{\mathrm{d}y} \left(\frac{1}{a_0^2} \right), \quad \frac{\mathrm{d}r}{\mathrm{d}\tau} = 0,$$

where $r = \partial \phi / \partial \tau$. Show that in $t > 0$ the projections of the rays through the origin are normal to the projections of the cross-sections $t =$ constant of the ray cones through the origin.

When $a_0^2 = 1/(1 + y)$, for $y > -1$, show that the ray cone through the origin is

$$x = p\tau, \quad y = \nu\tau + \frac{1}{4} r^2 \tau^2, \quad t = -r\tau - \frac{1}{2} r \nu \tau^2 - \frac{1}{12} r^3 \tau^3,$$

where

$$p^2 + \left(\nu + \frac{1}{2} r^2 \tau \right)^2 = (1 + y) r^2.$$

*8.14. Suppose ψ satisfies the modified Helmholtz equation

$$\nabla^2 \psi - k^2 \psi = 0$$

in the rectangle $-1 < x < 1$, $-a < y < a$, with $\psi = 1$ on the boundary. Show that the solution is

$$\psi = \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{n + 1/2}$$

$$\times \left(\frac{\cos((n + 1/2)\pi x) \cosh \alpha_n y}{\cosh \alpha_n a} + \frac{\cos((n + 1/2)\pi y / a) \cosh \beta_n x}{\cosh \beta_n} \right),$$

where $\alpha_n^2 = (n+1/2)^2\pi^2 + k^2$ and $\beta_n^2 = (n+1/2)^2\pi^2/a^2 + k^2$. Show also that

$$\psi = 2e^{-k}\cosh kx + 2e^{-ka}\cosh ky$$

satisfies the equation, that it is the limit of the exact solution as $k \to \infty$, and that it approximately satisfies the boundary conditions except near the corners. If this solution is written as $Ae^{-ku(x,y)}$, show that, to lowest order,

$$u \sim -(x+1), \quad x-1, \quad -(y+a), \quad y-a$$

near

$$x=-1, \quad x=1, \quad y=-a, \quad y=a,$$

respectively, and compare this with the answer to Exercise 8.1.

Remark. Near the ridge line $x=y$, ψ is approximately $e^{-kx}+e^{-ky}$, which suggests that across a ridge line u is continuous, tending to x on one side and y on the other, but ∇u has a jump discontinuity; specifically,

$$\left.\frac{\partial u}{\partial n}\right|_{-} = -\left.\frac{\partial u}{\partial n}\right|_{+}.$$

*8.15. Small-amplitude waves created by a ship moving along the x axis with speed $-V$ on an ocean $z<0$ are modelled by

$$\nabla^2\phi = 0 \quad \text{for } z<0,$$

with

$$\frac{\partial\phi}{\partial z} = \frac{\partial\eta}{\partial t}, \quad \frac{\partial\phi}{\partial t} + g\eta = 0 \quad \text{on } z=0,$$

where $\phi(x,y,z,t)$ is the velocity potential and $\eta(x,y,t)$ is the surface elevation. Show that, if $\xi = x+Vt$,

$$\frac{\partial\phi}{\partial z} = \frac{\partial\eta}{\partial t} + V\frac{\partial\eta}{\partial\xi}, \quad \frac{\partial\phi}{\partial t} + V\frac{\partial\phi}{\partial\xi} + g\eta = 0 \quad \text{on } z=0.$$

Now let

$$\eta = \Re\left(\eta_0 e^{i(-\omega t + k_1\xi + k_2 y)}\right).$$

Show that

$$\phi = \Re\, e^{i(-\omega t + k_1\xi + k_2 y) + z\sqrt{k_1^2+k_2^2}}$$

and that

$$g^2(k_1^2+k_2^2) = (-\omega + Vk_1)^4.$$

*8.16. In an optimal control problem for $x(t)$, it is desired to choose $u(t)$ to minimise a cost $\int_0^1 L(x,u)\,\mathrm{d}t$ when x evolves according to $\dot{x} = f(x,u)$. Assuming sufficient differentiability with respect to x and u, show that

$$\frac{\partial L(x,\dot{x})}{\partial \dot{x}} = \frac{\partial L}{\partial u} \Big/ \frac{\partial f}{\partial u} = p, \quad \text{say,}$$

and

$$\frac{\partial u(x,\dot{x})}{\partial \dot{x}} = 1 \Big/ \frac{\partial f}{\partial u}, \quad \frac{\partial u(x,\dot{x})}{\partial x} = -\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial u}.$$

Deduce from the Euler–Lagrange equation that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L(x,\dot{x})}{\partial \dot{x}} - \frac{\partial L(x,\dot{x})}{\partial x} &= \frac{\mathrm{d}p}{\mathrm{d}t} - \frac{\partial L}{\partial x} + \frac{\partial L}{\partial u}\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial u} \\ &= \dot{p} - \frac{\partial L}{\partial x} + p\frac{\partial f}{\partial x} \\ &= 0.\end{aligned}$$

Show further that, if $H(x,p)$ is defined to be $-L + pf$, where L and f are now thought of as functions of x and p, then

$$\frac{\partial H(x,p)}{\partial x} = -\frac{\partial L}{\partial x} + p\frac{\partial f}{\partial x} + \left(-\frac{\partial L}{\partial u} + p\frac{\partial f}{\partial u}\right)\frac{\partial u(x,p)}{\partial x} = -\dot{p},$$

$$\frac{\partial H(x,p)}{\partial p} = f + \left(-\frac{\partial L}{\partial u} + p\frac{\partial f}{\partial u}\right)\frac{\partial u(x,p)}{\partial p} = \dot{x};$$

thus the system is optimally controlled by a solution of Charpit's equations. Show that if $x(0) = x_0$ is prescribed then the second boundary condition is $p(1) = 0$.

*8.17. (i) *Fermat's principle* states that light travelling from the origin to $\mathbf{x}$ in a medium in which the speed of light is $a(\mathbf{x})$ does so along a path which minimises the optical path length

$$u(\mathbf{x}) = \int_0^{\mathbf{x}} \frac{\mathrm{d}s}{a}$$

between the origin and $\mathbf{x}$. Writing the path as $\mathbf{x} = \mathbf{X}(t)$, so that

$$u(\mathbf{x}) = \int_0^T \frac{|\dot{\mathbf{X}}|}{a(\mathbf{X})}\,\mathrm{d}t, \qquad \text{where} \qquad \mathbf{X}(T) = \mathbf{x}, \quad \mathbf{X}(0) = \mathbf{0},$$

show that the Euler–Lagrange equations for $\mathbf{X}$ are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{\mathbf{X}}}{a|\dot{\mathbf{X}}|}\right) + \frac{|\dot{\mathbf{X}}|}{a^2}\nabla a = 0, \quad \dot{} = \frac{\mathrm{d}}{\mathrm{d}t}.$$

Show further that, when τ is defined by $\mathrm{d}\tau/\mathrm{d}t = a|\dot{\mathbf{X}}|$, these equations can be written as

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau} = -\frac{1}{a^3}\nabla\left(a(\mathbf{X})\right), \quad \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} = \mathbf{p}.$$

(ii) Show that $|\nabla u(\mathbf{X})| = 1/a(\mathbf{X})$ and that the equations above are Charpit's equations for this partial differential equation.

(iii) Show that Charpit's equations are those for a Hamiltonian system in which
$$H(\mathbf{X}, \mathbf{p}) = \frac{1}{2}|\mathbf{p}|^2 - \frac{1}{2a^2}.$$

Remarks.

(a) By taking $1/a$ to be a suitable curvature, this idea can be used to reduce the problem of finding geodesics on a curved surface to the solution of a Hamilton–Jacobi equation.

(b) *In Lagrangian mechanics.*

(i) We take the minimum of the *action*
$$u(q_i) = \int_0^T L\left(Q_i(t), \dot{Q}_i(t)\right) \mathrm{d}t, \quad \dot{} = \frac{\mathrm{d}}{\mathrm{d}t},$$
over all paths $Q_i(t)$ with $Q_i(T) = q_i$, where L is the Lagrangian. Thus the Euler–Lagrange equations are
$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \frac{\partial L}{\partial Q_i}, \quad p_i = \frac{\partial L}{\partial \dot{Q}_i}.$$

(ii) By varying q_i slightly, we show that, for arbitrary δQ_i,
$$\sum \frac{\partial u}{\partial Q_i}(Q_i(T))\, \delta Q_i = L\left(Q_i(T), \dot{Q}_i(T)\right) \delta T,$$
and hence that
$$\sum \frac{\partial u}{\partial Q_i}\dot{Q}_i = L,$$
which means that
$$\frac{\partial u}{\partial Q_i} = \frac{\partial L}{\partial \dot{Q}_i} = p_i$$
from (i).

(iii) We note that, since $\mathrm{d}/\mathrm{d}T\left(u(Q_i) - \int_0^T L\,\mathrm{d}t\right) = 0$ (from the definition of u), u satisfies the Hamilton–Jacobi equation
$$\sum p_i \frac{\mathrm{d}Q_i}{\mathrm{d}t} - L \equiv H(Q_i, p_i) = \text{constant};$$
moreover, Charpit's equations are
$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i}, \quad \frac{\mathrm{d}Q_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i},$$
which are just Hamilton's equations.

(c) Note that the parameter t in Lagrangian mechanics is both the parameter along the characteristics and real time. In the Fermat model $\mathrm{d}\tau = a\,\mathrm{d}s$ and τ must be thought of as a weighted arc length along a light ray.

*9 Miscellaneous topics

9.1 Introduction

Our discussion over the past eight chapters has covered much ground but only scratched the surface of what is known about many methods for special classes of pdes.[168] In this miscellany we will attempt to take the reader further into some of the 'tricks of the trade'. However, before going into detail, we will make some general remarks about some areas of mathematics that are not only germane to the classical theory of pdes, but also so large and important as to warrant their own text books; yet we have scarcely mentioned them thus far.

The first concerns the study of *inverse problems*, some of which go under the name of *parameter identification*. This is a subject whose philosophy is almost orthogonal to ours in that it starts with observations about a phenomenon that may be modelled by some pde, and then asks 'what are the coefficients and boundary data for the relevant pde?' This is a very difficult procedure for the following reasons.

1. It is essentially nonlinear; even if a pde is linear, its solutions rarely depend linearly on its coefficients.
2. It can so easily be ill-posed. For example, suppose that the gravitational potential ϕ produced by a finite two-dimensional body of unknown constant density ρ satisfies the Poisson equation

$$\nabla^2\phi = -\rho,$$

and we can make measurements about this potential on a circle $r = R$, enclosing but not touching the body. What observations would be enough to determine its shape? Clearly, even a knowledge that $\partial\phi/\partial r = g$ is constant on $r = R$ does not determine ϕ, because any circular body with radius $\sqrt{2Rg/\rho}$ is possible.[169]

Even more difficult challenges are provided by the determination of unknown spatially-varying coefficients such as thermal conductivity. When, as is often the case, the measurements lead to an ill-posed *underdetermined* problem for the unknowns, there is clearly great scope for finding what is likely to be the 'best guess' by using various kinds of regularisation; see [18].

[168]In this chapter only, we will abbreviate 'partial differential equations' to 'pdes'.

[169]In fact, non-circular bodies are also possible in this case; examples can be constructed using the Hele-Shaw model of Chapter 7.

In view of these difficulties we will not take this discussion much further, except to note those aspects of the theory that we have developed in the previous chapters that might be of especial value in inverse problems. One commonly used technique for elliptic equations is to prescribe Dirichlet data and observe the resulting Neumann data, or vice versa. The relationship between these data is a generalisation of the Dirichlet-to-Neumann map (5.65) and (5.68) and contains information about the parameters in the pde away from the boundary; hence it can be of some help in turning parameter identification problems into integral equations. In another vein, the spectrum of a pde problem involving an eigenvalue as a parameter may easily be measured in practice by observing or hearing its 'normal modes' of vibration; this then poses the problem of reconstructing the appropriate Helmholtz equation (or whatever it might be) from the spectrum. This is a notoriously difficult problem, pioneered by a famous paper entitled 'Can you hear the shape of a drum?',[170] and it transpires that knowledge of the spectrum of, say, the Dirichlet problem alone is not sufficient to find the shape. To explain this and its importance to modern developments in radar and tomography requires far more knowledge of spectral theory than we were able to present in Chapter 5. This also applies to the more practically important problem of finding the shape or properties of a *scatterer* that is being insonified or irradiated by some incident wave field, as in §8.2.3. Indeed, the one theory we have encountered that is invaluable in such 'inverse scattering' problems is that of geometric optics. As long as the scattering body is much larger than the incident wavelength, the fact that the fields scattered by arbitrary bodies can be more or less drawn by hand means that it is pretty easy to build up intuition concerning the causes of any given scattered field. But the precise mathematical theory at arbitrary wavelength is much more difficult; some progress can be made in one space dimension, as we shall see at the end of this chapter.

While inverse problems can easily lead to underdetermined systems, in which the model is lacking information, the opposite phenomenon of *overdetermined systems* can also occur. In particular, such systems can arise where the ideas of group symmetry are being used, as in §6.5. There we saw that the invariance of a pde under a certain Lie group (or generalisation thereof) can only be assured if the functions defining the group are such that all the coefficients in the transformed pde coincide with those of the original. This inevitably leads to an enormous system of pdes for the defining functions, usually much larger in dimension than the number of defining functions.

An account of overdetermined systems can be found in [10]; here we simply note that the basic question of whether such a system has any solution at all[171] can be approached systematically by repeated cross-differentiation. The idea is to hit the system with more and more partial derivatives with respect to all the independent variables, and use the equality of mixed derivatives until either there is a contradiction or any further differentiation leads to equations which are

[170]Kac, M. (1966). 'Can you hear the shape of a drum?', *Amer. Math. Monthly*, **73**(4), 1–23.

[171]Compare, for example, the systems (i) $\partial u/\partial x = \partial u/\partial y = 0$ and (ii) $\partial u/\partial x = 0$, $\partial u/\partial y = x$ for the single function $u(x, y)$.

automatically satisfied. Of course, this is a tedious procedure and much time can be saved by using symbolic manipulators, which are now often incorporated into packages that search systematically for group symmetries.

Another area that has gone almost unnoticed in this book concerns the study of *stochastic pdes*, which is an increasingly important area as far as applications are concerned. One reason for this is the demand for risk quantification in areas of human activity ranging from insurance and finance to management and social policy. Unfortunately, the mastery of the subject makes two demands on the student, namely a good knowledge of stochastic processes in general as well as the development of the calculus needed for continuous-time random processes. In §§1.1 and 6.1 we have given a very crude description of some problems in this area, but we overlooked both of the issues just mentioned; if we were to take the subject any further we would immediately have to plunge more deeply into these topics. Hence we refer the reader to [31].

Our final apologia concerns the lack of discussion of pdes with non-smooth coefficients. Although we have placed much emphasis on non-smooth boundaries and boundary data, our only discontinuous-coefficient examples were the oxygen consumption model of §7.1.2 and the enthalpy formulation of the Stefan problem of §7.4.2. It is only too easy to assume that, if a pde has discontinuous coefficients, then the solution can be synthesised by piecing together smooth functions on either side of some interface. Such an intuitive approach not only relies on the existence of an interface, but also on its smoothness, and we recall the unexpected mushy regions that can occur in Stefan problems.[172] The whole question of how general pdes should be interpreted at places where the coefficients are discontinuous is even more delicate than those arising in the theory of shocks for conservation laws.

This chapter cannot attempt to compensate for the lack of coverage of these aspects of pdes. Instead, we will end the book by making some observations about some models and methods that do not fall directly under the headings of Chapters 1–8, but which illustrate the power and limitations of some of the ideas in those chapters. Inevitably. all the following sections are open-ended.

9.2 Linear systems revisited

We recall that our principal entrée into the theory of applied pdes was via quasilinear systems of the type $\sum_{i=1}^{m} \mathbf{A}_i\, \partial \mathbf{u}/\partial x_i = \mathbf{c}$ which, as suggested in the Introduction, cover all pdes with sufficiently smooth coefficients and with the coefficient of the relevant highest derivative being non-zero. We now reconsider some aspects of such systems in the light of what we have learned in Chapters 4–6. Little new general methodology emerges, but several special methods are available for particular problems.

Not surprisingly, for linear systems, the ideas of using Green's and Riemann functions that were so successful in Chapters 4–6 can be generalised, but this can only be done at some technical cost, as shown below.

[172] Further difficulties could arise in other practical problems, for example, if frictional effects were modelled by terms whose sign depends on the direction of motion.

9.2.1 Linear systems: Green's functions

First let us recall the situation for systems of ordinary differential equations (odes). The 'Cauchy problem' for $\mathbf{x}(t)$ such that

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{A}(t)\mathbf{x} = \mathbf{b}(t) \tag{9.1}$$

is to find $\mathbf{x}$ in $t > 0$, with

$$\mathbf{x}(0) = \mathbf{x}_0$$

prescribed. This can be solved in principle by finding the *Green's matrix* $\mathbf{G}(t, \tau)$ such that

$$-\frac{\mathrm{d}\mathbf{G}}{\mathrm{d}t} - \mathbf{G}\mathbf{A} = \mathbf{0}, \tag{9.2}$$

with

$$\mathbf{G} = \mathbf{I} \quad \text{at } t = \tau, \tag{9.3}$$

where $\mathbf{I}$ is the identity.[173] This formulation for $\mathbf{G}$ can easily be seen to be equivalent to

$$-\frac{\mathrm{d}\mathbf{G}}{\mathrm{d}t} - \mathbf{G}\mathbf{A} = \delta(t-\tau)\mathbf{I}, \tag{9.4}$$

with

$$\mathbf{G}(T, \tau) = \mathbf{0} \quad \text{for some } T > \tau, \tag{9.5}$$

which is analogous to the Riemann function formulation; equation (9.4) simply means that

$$\mathbf{G}(\tau - 0, \tau) - \mathbf{G}(\tau + 0, \tau) = \mathbf{I},$$

and hence

$$\mathbf{G}(\tau - 0, \tau) = \mathbf{I}.$$

With either definition of $\mathbf{G}$, we can proceed by premultiplying (9.1) by $\mathbf{G}$ and post-multiplying (9.2) by $\mathbf{x}$, as was done repeatedly in Chapters 4–6, and subtracting, to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{G}\mathbf{x}) = \mathbf{G}\mathbf{b}, \tag{9.6}$$

and hence

$$\mathbf{x}(\tau) = \mathbf{G}(0, \tau)\mathbf{x}(0) + \int_0^T \mathbf{G}(t, \tau)\mathbf{b}(t)\,\mathrm{d}t \tag{9.7}$$

since $\mathbf{G}(\tau, \tau) = \mathbf{I}$. Note that, if $\mathbf{A}$ = constant, so that[174] $\mathbf{G} = \mathrm{e}^{-\mathbf{A}(t-\tau)}$ for $t \leqslant \tau$, and $\mathbf{b}$ tends to a constant vector as $\tau \to \infty$, then it is easy to see that, as $\tau \to \infty$, (9.7) gives

$$\mathbf{x} \to -\mathbf{A}^{-1}\mathbf{b}$$

just as long as the eigenvalues of $\mathbf{A}$ have negative real parts.

[173] $\mathbf{G}$ is simply related to the so-called fundamental matrix that is often used in texts on odes [11].

[174] We define the matrix exponential by

$$\mathrm{e}^{t\mathbf{A}} = \sum_{n=0}^{\infty} \frac{t^n \mathbf{A}^n}{n!}.$$

Two-point boundary value problems for (9.1) can also be solved in this way. Suppose, for example, that we have the 'Dirichlet' problem when $\mathbf{x} = (x_1, \dots, x_{2n})^{\mathsf{T}}$, with $(x_1, \dots, x_n)^{\mathsf{T}} = \mathbf{0}$ at $t = 0$ and $t = 1$; then

$$\mathbf{x}(\tau) = \int_0^1 \mathbf{G}(t, \tau)\mathbf{b}(t)\,\mathrm{d}t, \tag{9.8}$$

where, to satisfy the boundary conditions, $\mathbf{G}$ is such that its last n columns vanish at $t = 0$ and $t = 1$. We can define $\mathbf{G}$ either by

$$-\frac{\mathrm{d}\mathbf{G}}{\mathrm{d}t} - \mathbf{G}\mathbf{A} = \delta(t - \tau)\mathbf{I} \tag{9.9}$$

or by (9.2) and

$$\left[\mathbf{G}\right]_{\tau-0}^{\tau+0} = -\mathbf{I}. \tag{9.10}$$

It is quite easy to relate (9.7) and (9.8) to the Green's function solutions of the scalar equations that we described in §§4.2.1 and 5.5.1, respectively. Indeed, if we were to write any of the scalar linear equations considered in Chapters 4–6 as first-order systems, as we always could, we would find that the Green's functions for these problems were simply appropriate entries in the Green's matrix defined by generalisations of (9.2) and (9.3) or (9.9) and (9.10). However, unless $\mathbf{A}$ is a constant matrix, it is unlikely that (9.1) could be recast as a scalar equation and thus, in general, the method described above is helpful. The reason for the added complexity of a matrix rather than a scalar Green's function is that, in order to solve a system, we need to know how each component of $\mathbf{x}$ responds to being driven by each and every component of $\mathbf{b}$.

Bearing these points in mind, let us now turn to the Cauchy problem for the linear hyperbolic system

$$\sum_{i=1}^{m} \mathbf{A}_i \frac{\partial \mathbf{u}}{\partial x_i} + \mathbf{B}\mathbf{u} = \mathbf{f} \quad \text{for } x_1 > 0, \tag{9.11}$$

with

$$\mathbf{u} = \mathbf{0} \quad \text{on, say, } x_1 = 0; \tag{9.12}$$

we assume that $x_1 = 0$ is not a characteristic as defined in (2.53). To solve this problem formally, we define a Green's (or Riemann) matrix as the solution of

$$-\sum_{i=1}^{m} \frac{\partial}{\partial x_i}(\mathbf{G}\mathbf{A}_i) + \mathbf{G}\mathbf{B} = \delta(\mathbf{x} - \boldsymbol{\xi})\mathbf{I}, \tag{9.13}$$

the delta function with vector argument being defined as in §4.2, and with $\mathbf{G}$ vanishing on a surface analogous to Γ in Fig. 4.1. Then, proceeding as usual,

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{x_1>0}\sum_{i=1}^{m}\frac{\partial(\mathbf{G}\mathbf{A}_i\mathbf{u})}{\partial x_i}\,\mathrm{d}\mathbf{x} = \mathbf{u}(\boldsymbol{\xi}) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{x_1>0}\mathbf{G}\mathbf{f}\,\mathrm{d}\mathbf{x}$$
$$= \mathbf{0}. \tag{9.14}$$

If, instead, we specify the initial data $\mathbf{u} = \mathbf{u}_0(\mathbf{x}')$ for $x_1 = 0$, where $\mathbf{x}' = (x_2, \ldots, x_m)^\mathsf{T}$, and $\mathbf{f} = \mathbf{0}$, our integration provides the standard formula

$$\mathbf{u}(\boldsymbol{\xi}) = \int_{x_1=0} \mathbf{G}\mathbf{A}_1\mathbf{u}_0\,d\mathbf{x}'. \tag{9.15}$$

What is not standard is the fact that, for m-dimensional problems, even for $x_1 < \xi_1$, $\mathbf{G}$ is a distribution with 'mass' concentrated on the $(m-1)$-dimensional characteristic surfaces through $\mathbf{x} = \boldsymbol{\xi}$. Hence the integrals in (9.14) and (9.15) include contributions from these surfaces and their intersection with $x_1 = 0$, as well as from the region inside. This is a crude generalisation of the discussion about propagation of discontinuities for the Riemann functions that we gave at the end of §4.2.2.

An exactly similar calculation can be carried out for the Dirichlet problem for an elliptic system of the form (9.11) with, say,

$$\mathbf{u} = (u_1, \ldots, u_{2n})^\mathsf{T}, \quad \text{and} \quad (u_1, \ldots, u_n)^\mathsf{T} = \mathbf{0}$$

on the boundary ∂D of a closed region D. We again define $\mathbf{G}$ to satisfy (9.13), but now its last n columns vanish on ∂D. The asymptotic behaviour of $\mathbf{G}$ near $\mathbf{x} = \boldsymbol{\xi}$ is now less easy to discern than it is for the ode case, as can be illustrated with the following model.

9.2.2 Linear elasticity

In §4.7.1, we have already encountered the famous Navier–Lamé equations of linear elasticity. In the case of static equilibrium in three dimensions, they are

$$\mathcal{L}\mathbf{u} \equiv \mu\nabla^2\mathbf{u} + (\lambda + \mu)\nabla\nabla\cdot\mathbf{u} = -\mathbf{f}(\mathbf{x}), \tag{9.16}$$

where $\mathbf{u}$ is the displacement, $\mathbf{f}$ is the body force per unit volume and λ and μ are the so-called *Lamé constants* which characterise the material.

In view of the identities

$$\mathbf{v}\cdot\nabla(\nabla\cdot\mathbf{u}) - \mathbf{u}\cdot\nabla(\nabla\cdot\mathbf{v}) = \nabla\cdot(\mathbf{v}\nabla\cdot\mathbf{u} - \mathbf{u}\nabla\cdot\mathbf{v})$$

and

$$\mathbf{v}\cdot\nabla\wedge(\nabla\wedge\mathbf{u}) - \mathbf{u}\cdot\nabla\wedge(\nabla\wedge\mathbf{v}) = \nabla\cdot((\nabla\wedge\mathbf{u})\wedge\mathbf{v} - (\nabla\wedge\mathbf{v})\wedge\mathbf{u}),$$

it is easy to see that, when suitable boundary conditions are prescribed for $\mathbf{u}$, the operator $\mathcal{L}$ is self-adjoint.

There are two remarks to be made before we start. First, the system is second order rather than first; although it could, with labour, be recast in the form (9.11), there is no reason why any of the discussion above should be confined only to linear systems of first-order pdes, and we will thus proceed directly with (9.16). Secondly, (9.16) has constant coefficients and hence could, by cross-differentiation, be reduced to a scalar equation for, say, any of the components of $\mathbf{u}$. This is also an unnecessarily tiresome procedure because, although the outcome is relatively simple (see Exercise 9.4), the physical significance of the Green's matrix is soon obscured.

For simplicity, let us just consider the solution of (9.16) in the case when the elastic continuum extends to infinity in all directions, and let us assume that $|\mathbf{u}|$ grows no faster than $|\mathbf{x}|$ at infinity and that the strain components $\partial u_i/\partial x_j + \partial u_j/\partial x_i$ decay at least as fast as $O(1/|\mathbf{x}|^3)$.[175] We note that the rigid body displacement $\mathbf{u} = \mathbf{c}$, where $\mathbf{c}$ is constant, satisfies $\mathcal{L}\mathbf{u} = \mathbf{0}$ and the decay conditions above, and hence the Fredholm Alternative implies that[176]

$$\int \mathbf{f}(\mathbf{x}) \cdot \mathbf{c}\, d\mathbf{x} = 0$$

for all $\mathbf{c}$, so that

$$\int \mathbf{f}(\mathbf{x})\, d\mathbf{x} = \mathbf{0}.$$

Also $\mathcal{L}\mathbf{u} = \mathbf{0}$, with the decay conditions, is satisfied by $\mathbf{u} = \boldsymbol{\omega} \wedge \mathbf{x}$ for constant $\boldsymbol{\omega}$, which corresponds to a rigid body rotation (since we are only considering small displacements $\mathbf{u}$, a rotation is represented by a vector product). Hence

$$\int \mathbf{f}(\mathbf{x}) \cdot (\boldsymbol{\omega} \wedge \mathbf{x})\, d\mathbf{x} = 0$$

for all $\boldsymbol{\omega}$, so that

$$\int \mathbf{x} \wedge \mathbf{f}(\mathbf{x})\, d\mathbf{x} = \mathbf{0}.$$

Physically, these are the conditions that $\mathbf{f}$ exerts zero net force and moment on the elastic continuum.

Now, again using the fact that $\mathcal{L}$ is self-adjoint, we are guided by (9.13) to simply define the Green's matrix $\mathbf{G}$ to be the solution of

$$\mathcal{L}\mathbf{G}(\mathbf{x} - \boldsymbol{\xi}) = -\delta(\mathbf{x} - \boldsymbol{\xi})\mathbf{I}, \tag{9.17}$$

with

$$\mathbf{G}(\mathbf{x} - \boldsymbol{\xi}) \to \mathbf{0} \quad \text{as } |\mathbf{x}| \to \infty. \tag{9.18}$$

each side being interpreted entry by entry. It is easy to see that $\mathbf{G}$ is symmetric and depends only on $\mathbf{x} - \boldsymbol{\xi}$; our usual procedure gives at once that

$$\int \left((\mathcal{L}\mathbf{G})\,\mathbf{u} - \mathbf{G}\mathcal{L}\mathbf{u}\right) d\mathbf{x} = -\int \big(\delta(\mathbf{x} - \boldsymbol{\xi})\mathbf{u}(\mathbf{x}) - \mathbf{G}(\mathbf{x} - \boldsymbol{\xi})\mathbf{f}(\mathbf{x})\big)\, d\mathbf{x}, \tag{9.19}$$

and hence, since the left-hand side vanishes, that

$$\mathbf{u}(\boldsymbol{\xi}) = \int \mathbf{G}(\mathbf{x} - \boldsymbol{\xi})\mathbf{f}(\mathbf{x})\, d\mathbf{x}.$$

It is interesting to note that, were we to carry out this calculation *ab initio* without relying on the validity of (9.19) for distributions (cf. the comments after (4.13)),

[175]These conditions ensure that the elastic strain energy is finite.

[176]For the rest of this section we denote integrals over the whole of three-dimensional space by $\int \cdots d\mathbf{x}$.

then the application of Green's theorem yields a contribution from a small sphere centred at $\boldsymbol{\xi}$ that can be interpreted in terms of the forces applied to that sphere. Thus $\mathbf{G}$ can be interpreted physically in terms of the response to arbitrarily directed point forces applied to the medium at $\mathbf{x} = \boldsymbol{\xi}$ (and hence we do not expect its entries to decay as rapidly at infinity as we insisted for $\mathbf{u}$ above).

In order to find $\mathbf{G}$ explicitly, we simply look at the Fourier transform of (9.17). Defining

$$\hat{\mathbf{G}}(\mathbf{k}, \boldsymbol{\xi}) = \int \mathbf{G}(\mathbf{x} - \boldsymbol{\xi}) e^{i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}, \tag{9.20}$$

where the transform variable $\mathbf{k}$ is now a vector, it is easy to see that[177]

$$\hat{\mathbf{G}}(\mathbf{k}, \boldsymbol{\xi}) = \left(\frac{1}{\mu |\mathbf{k}|^2} \mathbf{I} - \frac{\lambda + \mu}{\mu(\lambda + 2\mu)} \frac{1}{|\mathbf{k}|^4} \mathbf{k}\mathbf{k}^{\top} \right) e^{i\mathbf{k}\cdot\boldsymbol{\xi}}. \tag{9.21}$$

Some care is needed with the Fourier inversion of this singular function (see Exercise 9.5), but it transpires that the entries in $\mathbf{G}$ are

$$\mathbf{G}_{ij}(\mathbf{x} - \boldsymbol{\xi}) = \frac{\lambda + \mu}{8\pi\mu(\lambda + 2\mu)} \frac{1}{|\mathbf{x} - \boldsymbol{\xi}|} \left(\frac{\lambda + 3\mu}{\lambda + \mu} \delta_{ij} + \frac{(x_i - \xi_i)(x_j - \xi_j)}{|\mathbf{x} - \boldsymbol{\xi}|^2} \right). \tag{9.22}$$

The presence of off-diagonal elements would have been difficult to spot directly from (9.17), and they illustrate the statement made above, that each component of $\mathbf{f}$ influences all the components of $\mathbf{u}$. The columns of $\mathbf{G}$ are precisely the above-mentioned displacement vectors associated with the 'point forces' along the coordinate axes.

When boundaries are present, (9.22) still describes the local behaviour of $\mathbf{G}$ near $\mathbf{x} = \boldsymbol{\xi}$, but, not surprisingly, the application of the concept of images is now much more complicated than in §5.6.

9.2.3 Linear inviscid hydrodynamics

We now consider what can happen to singular solutions of systems that have a real characteristic but are not hyperbolic, thereby again illustrating the surprising way in which the components of $\mathbf{f}$ in (9.11) can influence the components of $\mathbf{u}$. We will study a very simple model of inviscid hydrodynamics which shows how the fundamental concepts of lift and drag on a body moving through a stationary liquid can be understood in terms of distributional solutions of a system of linear pdes. The model is a truncated version of the incompressible Euler equations, to which (2.3)–(2.7) are related, and to which we will refer again in §9.4.4. When the fluid is being driven by a localised force that is so weak that only linear terms need be retained, these equations take the dimensionless form[178]

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla p = \mathbf{f}, \quad \nabla \cdot \mathbf{u} = 0,$$

with $\mathbf{u} = \mathbf{0}$ at $t = 0$, and we now study the response to various forms of $\mathbf{f}$ which might describe different kinds of small bodies moving through the fluid.

[177] By $\mathbf{k}\mathbf{k}^{\top}$ we mean the matrix $(k_i k_j)$.

[178] In this chapter only, we revert to the conventional use of $\mathbf{u}$ as the fluid velocity in hydrodynamics.

In all cases we assume the fluid extends to infinity in all directions and is at rest there. Also, we expect that much of the flow is irrotational because, in regimes where $\mathbf{f} = \mathbf{0}$, the vorticity $\nabla \wedge \mathbf{u}$ is independent of time. Hence, if the vorticity is zero initially, then we expect it to stay zero, but we will encounter some surprises in this respect.

A lift force in two dimensions

Consider a small lifting body, such as an aerofoil moving along the x axis. We assume that the lift force $L\mathbf{j}$, in the y direction, is prescribed, along with the aerofoil position $X(t)\mathbf{i}$. When we take axes moving with the aerofoil, we expect $\mathbf{u}$ and p to satisfy

$$\frac{\partial \mathbf{u}}{\partial t} - \dot{X}(t)\frac{\partial \mathbf{u}}{\partial \xi} + \nabla p = L\delta(\xi)\delta(y)\mathbf{j}, \quad \nabla \cdot \mathbf{u} = 0, \tag{9.23}$$

where $\xi = x - X(t)$. When L and $\dot{X}$ are constant, so that we can set $\partial \mathbf{u}/\partial t = \mathbf{0}$, we find that the Fourier transform

$$\hat{\mathbf{u}}(k, y) = \int_{\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{u}e^{ik_1\xi + ik_2 y}\,d\xi\,dy \tag{9.24}$$

is given by

$$\hat{\mathbf{u}} = \frac{L}{\dot{X}}\frac{i}{k_1^2 + k_2^2}(k_2\mathbf{i} - k_1\mathbf{j}),$$

with

$$\hat{p} = \frac{ik_2 L}{k_1^2 + k_2^2}.$$

It is a simple matter to invert these transforms using the fact that the transform of $\log r$ is $-2\pi/(k_1^2 + k_2^2)$ because $\nabla^2 \log r = 2\pi\delta(x)\delta(y)$; the answer is

$$\mathbf{u} = \frac{L}{2\pi\dot{X}}\frac{1}{\xi^2 + y^2}(y\mathbf{i} - \xi\mathbf{j}), \quad p = \frac{Ly}{2\pi(\xi^2 + y^2)}, \tag{9.25}$$

which is the velocity field of a vortex centred at $x = X(t)$, $y = 0$. Hence a small moving aerofoil can be identified with such a vortex.

A quite different result sometimes emerges when we allow the velocity and the lift on the aerofoil to vary. As long as $L = \Gamma\dot{X}$, where Γ is constant, we find

$$\frac{d}{dt}\left(\hat{\mathbf{u}}e^{ik_1 X}\right) = \frac{\Gamma\dot{X}e^{ik_1 X}}{k_1^2 + k_2^2}(-k_2\mathbf{i} + k_1\mathbf{j}), \tag{9.26}$$

and hence we retrieve (9.25). However, as soon as $L \neq \Gamma\dot{X}$, (9.26) cannot be integrated directly and $\hat{\mathbf{u}}$ acquires a 'history'. For example, if $L = \Gamma\dot{X}H(t)$, where H is the Heaviside function, (9.25) is replaced by

$$\mathbf{u} = \frac{\Gamma}{2\pi}\left(\frac{y\mathbf{i} - \xi\mathbf{j}}{\xi^2 + y^2} - \frac{y\mathbf{i} - x\mathbf{j}}{x^2 + y^2}\right), \tag{9.27}$$

which shows that a 'starting vortex' is shed at the origin $x = y = 0$.[179] Our expectation that the flow would be irrotational away from the aerofoil is unjustified in this case; the single real characteristic of (9.23) has propagated information about the motion of the aerofoil.

Drag in two dimensions

This is an interesting problem to pose because it is well known that d'Alembert's paradox [27] precludes the existence of a drag force on a body moving with constant velocity in an irrotational flow. However, if we attempt to model a drag force D by naively replacing (9.23) by

$$\frac{\partial \mathbf{u}}{\partial t} - \dot{X}\frac{\partial \mathbf{u}}{\partial \xi} + \nabla p = D\delta(\xi)\delta(y)\mathbf{i}, \tag{9.28}$$

we find that, when D and $\dot{X}$ are constant and $\partial \mathbf{u}/\partial t = \mathbf{0}$,

$$\hat{\mathbf{u}} = \frac{-\mathrm{i}D}{\dot{X}(k_1^2 + k_2^2)}\left(\frac{k_2^2}{k_1}\mathbf{i} - k_2\mathbf{j}\right), \quad \hat{p} = \frac{\mathrm{i}Dk_1}{k_1^2 + k_2^2}, \tag{9.29}$$

and inversion leads to the appearance of a delta function in the components of $\mathbf{u}$ as we approach the drag-producing body at $\xi = y = 0$. In fact, it is well known that drag on an accelerating body, usually referred to as ***added mass***, must be modelled by

$$\frac{\partial \mathbf{u}}{\partial t} - \dot{X}\frac{\partial \mathbf{u}}{\partial \xi} + \nabla p = \mathbf{0}, \quad \nabla \cdot \mathbf{u} = M\delta'(\xi)\delta(y). \tag{9.30}$$

The last term is a 'mass dipole' and it leads to the velocity field

$$\mathbf{u} = \frac{M}{2\pi}\left(\frac{\partial^2}{\partial \xi^2}(\log r)\mathbf{i} + \frac{\partial^2}{\partial \xi \partial y}(\log r)\mathbf{j}\right), \quad r^2 = \xi^2 + y^2, \tag{9.31}$$

which differs from the inverse transform of (9.29) by the aforementioned delta function in $\mathbf{u}$. Hence (9.28) is not a good representation of a drag force. The correct representation (9.30) models the flow generated by a small non-lifting body for arbitrary $\dot{X}$ and the exact solution of the Euler equations reveals that M is proportional to $\ddot{X}$.

Three-dimensional flow

The situation becomes even more interesting in three dimensions. When a small aerofoil at $(Vt, 0, 0)$ is modelled by

$$-V\frac{\partial \mathbf{u}}{\partial \xi} + \nabla p = L\delta(\xi)\delta(y)\delta(z)\mathbf{j}, \quad \nabla \cdot \mathbf{u} = 0,$$

we find the three-dimensional Fourier transform

[179] When viscosity is taken into account, this vortex diffuses through the whole flow field, which is eventually steady and such that $\int \mathbf{u} \cdot d\mathbf{s}$ around the aerofoil is equal to Γ, which is called the circulation.

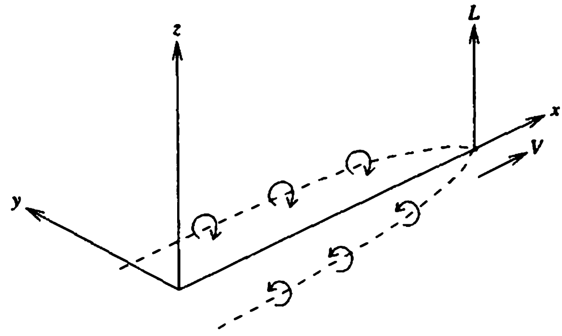


Fig. 9.1 Horseshoe vortex.

$$\hat{\mathbf{u}} = -\frac{\mathrm{i}Lk_2}{Vk_1(k_1^2 + k_2^2 + k_3^2)}\mathbf{k}.$$

Hence $\mathbf{u} = \nabla\phi$, where

$$\phi = \frac{L}{4\pi V}\frac{y}{y^2+z^2}\left(1 - \frac{x - Vt}{\left((x-Vt)^2 + y^2 + z^2\right)^{1/2}}\right),$$

so now, even in steady flow, a wake is shed along the degenerate characteristic $y = z = 0$, $x < Vt$. This is the famous 'horseshoe vortex' [26] (see Fig. 9.1).

Our approach can even shed light on the well-known phenomenon of vortex ring propagation, as produced, say, by a suitably exhaled puff of cigarette smoke. Using the technique described above, we can show that the solution of a steady axisymmetric flow modelled by

$$-V\frac{\partial \mathbf{u}}{\partial \xi} + \nabla p = D\delta(\xi)\delta'(y)\delta'(z)\mathbf{i}, \quad \nabla\cdot\mathbf{u} = 0$$

is

$$\mathbf{u} = \frac{D(2\xi^2 - y^2 - z^2, 3\xi y, 3\xi z)}{4\pi\left(\xi^2 + y^2 + z^2\right)^{5/2}} \quad \text{for } (\xi, y, z) \neq 0,$$

which is the far-field of a vortex ring (see [39] and (7.114)). This velocity field differs by a delta function at the origin from that produced by a mass dipole as in (9.30), but, more importantly, it shows that we can regard the motion produced by a smoke ring as equivalent to that generated by a suitable point force in the equations of motion.

9.2.4 Wave propagation and radiation conditions

For linear systems of pdes which model wave propagation, it is possible to proceed directly in the time domain by analysing the equation with three independent space variables and time t, say by using Riemann matrix ideas. However, we have often remarked that, if the coefficients in the system are independent of time, then it is easier and frequently of more practical interest to restrict attention to monochromatic waves in the frequency domain by writing the dependent variables as functions proportional to $e^{-i\omega t}$, say.[180] We did that for the scalar wave equation in §§5.1.5 and 8.1, and some of the ideas in those sections carry over to the vector case. However, many fascinating problems arise, such as the question of propagation through periodic media. As mentioned in §4.5.4, in one dimension this leads to the notions of pass and stop bands bounded by eigenvalues of a periodic Sturm–Liouville problem, but this framework relies on the well-developed Floquet theory for odes with periodic coefficients. It is much harder to understand multidimensional configurations of this kind because of the geometric complexity of the waves as they reflect from each periodic cell boundary. It is a great pity that we cannot discuss this further here because of the fundamental importance of wave propagation in crystal lattices.

Another issue that arises when we consider problems in the frequency domain for unbounded wave-bearing media that are uniform at large distances is the question of the radiation conditions that generalise the Sommerfeld condition (5.75) which was so vital in ensuring uniqueness for scalar problems. For example, for Maxwell's equations (see §4.7.2) we may suppose that the leading-order far-field radiation due to any bounded source of non-zero intensity decays with distance in proportion to e^{ikr}/r (or $e^{ikr}/\sqrt{r}$ in two dimensions), where $k = \omega/c$ is the wavenumber. Now all electromagnetic waves are transverse waves in the sense that the directions of the fields $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to the direction of propagation of any wave; this is a trivial consequence of (4.82) in the frequency domain, because, when we seek such solutions in which $\mathbf{E} = \mathbf{E}(\mathbf{k} \cdot \mathbf{r})$ and $\mathbf{H} = \mathbf{H}(\mathbf{k} \cdot \mathbf{r})$, both $\mathbf{E}$ and $\mathbf{H}$ are clearly perpendicular to $\mathbf{r}$. Hence we may write that, at large distances, where $\mathbf{k}$ and $\mathbf{r}$ are nearly parallel,

$$\mathbf{E} \sim \frac{e^{ikr}}{r}\mathbf{e}, \quad \mathbf{H} \sim \frac{e^{ikr}}{r}\mathbf{h},$$

where $\mathbf{e}$ and $\mathbf{h}$ are both azimuthal and are independent of $\mathbf{r}$ to lowest order. Then (4.82), suitably scaled, gives at once the radiation conditions

$$\mathbf{0} = \nabla \wedge \mathbf{E} - i\omega\mathbf{H} \sim i(\mathbf{k} \wedge \mathbf{e} - \omega\mathbf{h}) = o\left(\frac{1}{r}\right),$$
$$\mathbf{0} = \nabla \wedge \mathbf{H} + i\omega\mathbf{E} \sim i(\mathbf{k} \wedge \mathbf{h} + \omega\mathbf{e}) = o\left(\frac{1}{r}\right)$$

as $r \to \infty$.

Similarly, in elasticity, the wave equations (4.73) imply that, far from any bounded region of sources, $\mathbf{u}$ decays like $\mathbf{u}_0 e^{ikr}/r$, where again $\mathbf{k}$ is radial. As expected, we find that either

$$\mathbf{k} = \mathbf{k}_s, \qquad \text{where} \qquad \mathbf{u}_0 \cdot \mathbf{k}_s = 0 \quad \text{and} \quad \omega^2 = \frac{k_s^2 \mu}{\rho},$$

[180]Such waves can always be superimposed to generate solutions in the time domain, but this may be easier said than done.

corresponding to transverse shear waves (S-waves), or

$$\mathbf{k} = \mathbf{k}_p, \qquad \text{where} \qquad \mathbf{u}_0 \wedge \mathbf{k}_p = \mathbf{0} \quad \text{and} \quad \omega^2 = \frac{k_p^2(\lambda + 2\mu)}{\rho},$$

corresponding to longitudinal compressive waves (P-waves). Hence the radiation condition is that, at large distances,

$$\mathbf{u} \sim \mathbf{u}_s \frac{e^{ik_s r}}{r} + \mathbf{u}_p \frac{e^{ik_p r}}{r},$$

where $\mathbf{u}_s \cdot \mathbf{k}_s = 0$ and $\mathbf{u}_p \wedge \mathbf{k}_p = \mathbf{0}$.

9.3 Complex characteristics and classification by type

We have highlighted in many places the dramatic differences between ellipticity, parabolicity and hyperbolicity. One attempt that has been made to reconcile these two concepts is Garabedian's theory of *complex characteristics* [21]. The basic idea relies on the fact that the solution of an elliptic partial differential equation is usually an analytic function of the independent variables. Hence, for the elliptic equation

$$\frac{\partial^2\phi}{\partial\xi^2} + \frac{\partial^2\phi}{\partial y^2} = 0,$$

we could seek solutions analytic in $\zeta = \xi + i\eta$ and note that this implies that

$$\frac{\partial^2\phi}{\partial\xi^2} + \frac{\partial^2\phi}{\partial\eta^2} = 0.$$

Subtracting the two equations, we see that

$$\frac{\partial^2\phi}{\partial y^2} - \frac{\partial^2\phi}{\partial\eta^2} = 0,$$

which is a *hyperbolic* equation in (y, η). However, this procedure inevitably involves analytic continuation of the boundary data which, as we have already remarked, is a dangerous procedure, and it is only justified as long as no singularities are encountered in the continuation into $\eta \neq 0$ (see Exercise 9.7).

Armed with our knowledge of ray theory from Chapter 8, we can gain further insight into the pitfalls encountered in trying to unify ellipticity, parabolicity and hyperbolicity. Consider, for example, the following problems in $y > 0$:

$$\alpha\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0, \tag{9.32}$$

with

$$\phi(x, 0) = e^{-x^2/2\epsilon}, \quad \epsilon > 0, \tag{9.33}$$

and either

(i) $\qquad \alpha > 0 \quad$ and $\quad \phi \to 0 \quad$ as $y \to \infty$

or

(ii) $\qquad \alpha < 0 \qquad$ and $\qquad \dfrac{\partial \phi}{\partial y} + \sqrt{-\alpha}\dfrac{\partial \phi}{\partial x} = 0 \quad \text{on } y = 0.$

It is easy to see that the decay condition is sufficient to ensure uniqueness in the elliptic case (i); the Cauchy data, which is equivalent to saying that ϕ is a function of $x - \sqrt{-\alpha}y$, ensures uniqueness in case (ii). (When $\alpha = 0$, the 'parabolic' case, the only solution that is bounded at infinity is $\phi(x, y) = e^{-x^2/2\epsilon}$.)

Although the parameter ϵ can be scaled out of either problem, we might expect that sensible limits could be retrieved as $\epsilon \to 0$. Indeed, it is possible that in this limit ϕ would be closely related to the Green's function or the Riemann function, respectively, and we will see the precise relation later. Now a ray theory ansatz in which $\phi \sim Ae^{u/\epsilon}$ yields the eikonal equation

$$\alpha\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 = 0,$$

and Charpit's method then gives $u = -s^2/2$, where, in case (i),

$$x = s(1-t), \quad y = \pm\frac{ist}{\sqrt{\alpha}}, \quad u = -\frac{(x \pm iy\sqrt{\alpha})^2}{2},$$

and so there is no possibility of exponential decay as $y \to \infty$. In case (ii), however, we have

$$x = s(1-t), \quad y = -\frac{st}{\sqrt{-\alpha}}, \quad u = -\frac{(x - y\sqrt{-\alpha})^2}{2}.$$

Thus in case (ii) the characteristics are $x = s + y/\sqrt{-\alpha}$, and we have retrieved the hoped-for propagation of the data along the characteristic $x = y\sqrt{-\alpha}$, but case (i) has woefully failed to reveal anything like the singularity associated with the Green's function for (9.32). The ray ansatz has worked well when the equation is hyperbolic but leads to solutions with unbounded oscillations and growth in the elliptic case.

Now we can solve either case by Fourier transforms. The answer is

(i) $$\phi(x, y) = \frac{\sqrt{\epsilon}}{\pi\sqrt{2}}\int_{-\infty}^{\infty} e^{-\epsilon k^2/2 - \sqrt{\alpha}|k|y - ikx}\, dk, \tag{9.34}$$

(ii) $$\phi(x, y) = \frac{\sqrt{\epsilon}}{\pi\sqrt{2}}\int_{-\infty}^{\infty} e^{-\epsilon k^2/2 + i\sqrt{-\alpha}ky - ikx}\, dk; \tag{9.35}$$

in each case the coefficient of y in the exponent is determined uniquely by the data.

We now have to rely on some results from the theory of asymptotic expansions of integrals [22]. As $\epsilon \to 0$, (9.35) is always approximated by the saddle point contribution that emerges when we write $k = \kappa/\epsilon$ to obtain

$$\frac{1}{\pi\sqrt{2\epsilon}}\int_{-\infty}^{\infty} e^{-(\kappa^2/2 + i\sqrt{-\alpha}\kappa y - i\kappa x)/\epsilon}\, d\kappa.$$

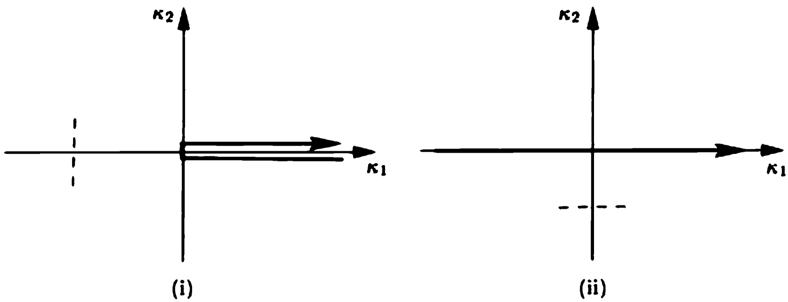


Fig. 9.2 Inversion contours for case (i) and case (ii); we have taken $x = 0$ in both cases. The steepest descents direction at the saddle is shown dashed.

The saddle point is at $\kappa = \mathrm{i}(x - y\sqrt{-\alpha})$, leading to a term $\mathrm{e}^{-\frac{1}{2}(x-y\sqrt{-\alpha})^2/\epsilon}$ in ϕ. However, the integral for case (i) must be written as

$$\Re\frac{1}{\pi\sqrt{2\epsilon}}\int_0^\infty \mathrm{e}^{-(\kappa^2/2+\sqrt{\alpha}\kappa y-\mathrm{i}\kappa x)/\epsilon}\,\mathrm{d}\kappa,$$

which is always approximated by its end-point contribution from near $\kappa = 0$, namely $\epsilon y\sqrt{\alpha}/(\alpha y^2 + x^2)$, which is proportional to the y derivative of the Green's function for (9.32) with Dirichlet data on $y = 0$.

Thus we see that the switch from elliptic to hyperbolic can be identified with the presence or absence of a saddle point on the inversion contour for a Fourier integral. In case (i), the inversion path cannot be deformed into a 'steepest descents' contour (see Fig. 9.2(i)), but in case (ii) it can (Fig. 9.2(ii)). Note that the function that emerges from the ray ansatz has no Fourier transform in the usual sense and is hence legislated against when we write down (9.34).

9.4 Quasilinear systems with one real characteristic

Annoyingly, practical problems often give rise to systems of pdes which, like that in §9.2.3, are neither elliptic nor hyperbolic in the sense of Chapter 3 and are often nonlinear as well. Hence they usually need to be treated individually on their merits and here we will simply list three examples.

9.4.1 Heat conduction with ohmic heating

In models of many electric devices, such as thermistors, the equations of conductive heat transfer must be coupled with those of electromagnetism via the resistance or 'ohmic' heating of the device. In one of the simplest configurations, the current density $\mathbf{j}$ is related to the electric potential ϕ by Ohm's law

$$\mathbf{j} = -\sigma(T)\nabla\phi, \tag{9.36}$$

where the electric conductivity σ is a positive function of the temperature T. In quasi-steady conditions, conservation of charge gives

$$\nabla \cdot \mathbf{j} = \nabla \cdot (-\sigma(T)\nabla\phi) = 0. \tag{9.37}$$

Meanwhile, the resistance heating generates a volumetric heat source $-\mathbf{j} \cdot \nabla\phi$ and so the energy equation in suitable units is

$$\frac{\partial T}{\partial t} = \nabla^2 T + \sigma(T)|\nabla\phi|^2. \tag{9.38}$$

Following §2.6, it is easy to check that the only real characteristic surfaces of (9.37) and (9.38) are t = constant but, as is often the case, this gives disappointingly little information about the structure of the solution. It simply suggests that the imposition of initial conditions on T and boundary conditions for T and ϕ might be sufficient to guarantee well-posedness, at least as long as σ is strictly positive. The only other obvious indicator is when there is enough symmetry (e.g. as in a one-dimensional problem) for (9.37) to be integrated explicitly and ϕ eliminated to yield a parabolic equation for T, which may be 'non-local' if the thermistor is coupled to an external circuit. In the steady case, however, in two dimensions the resulting elliptic system has the remarkable property that it is invariant under conformal maps, and hence, if the boundary conditions have the right form, many explicit solutions can be found (see Exercise 9.8).

9.4.2 Space charge

Another interesting electromagnetic phenomenon occurs when a field is generated that is strong enough to inject ions into the medium in which it is acting. This can happen in the air around a high-voltage DC transmission line; another example is the use of electrostatic precipitation to remove small particles from power-station emissions or to apply a coating of charged paint particles to a metal workpiece. Since all the current is carried by the mobile ions, it is reasonable to take the current density $\mathbf{j}$ to be the product of the average charge density ρ and the average particle velocity $\mathbf{v}$, so that

$$\mathbf{j} = \rho\mathbf{v}.$$

In many cases, $\mathbf{v}$ is found by balancing the electrostatic force on an ion or charged particle, which is proportional to $-\nabla\phi$, with a viscous drag force proportional to $\mathbf{v}$. Hence, in suitable units,

$$\mathbf{j} = -\rho\nabla\phi.$$

Now, conservation of charge gives

$$\frac{\partial \rho}{\partial t} - \nabla \cdot (\rho\nabla\phi) = 0, \tag{9.39}$$

whereas Maxwell's equations with $\epsilon = 1$ demand that ρ and ϕ are related by Poisson's equation

$$\nabla^2\phi = -\rho. \tag{9.40}$$

When we now follow §2.6, we find the normal cone again has only one real component and that it is a plane. The characteristic cone is thus the line $(-\nabla\phi, 1)$ in space–time. In a steady problem this gives considerable insight because it is easily seen that (9.39) reduces to

$$\nabla\phi\cdot\nabla\rho-\rho^2=\frac{\mathrm{d}\rho}{\mathrm{d}\tau}+\rho^2=0,$$

where $\mathrm{d}/\mathrm{d}\tau$ denotes differentiation along a characteristic. Moreover, in steady two-dimensional problems, there is a 'stream function' ψ such that

$$\rho\frac{\partial\phi}{\partial x}=\frac{\partial\psi}{\partial y},\quad \rho\frac{\partial\phi}{\partial y}=-\frac{\partial\psi}{\partial x},$$

and ψ is constant on a characteristic, which is thus orthogonal to the equipotentials. All these observations enable us to simplify the problem by transforming to hodograph variables ϕ and ψ. Some solutions are given in Exercise 9.9.

9.4.3 Fluid dynamics: the Navier–Stokes equations

It is very helpful to keep several of the ideas introduced above in mind when considering the far more important and difficult Navier–Stokes equations of hydrodynamics. This is probably the most intensively studied of all systems of nonlinear pdes and it is a synthesis of two different models described earlier in the book. For incompressible flow, the system is

$$\frac{\partial\mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}=-\nabla p+\frac{1}{R}\nabla^2\mathbf{u},\quad \nabla\cdot\mathbf{u}=0, \tag{9.41}$$

with appropriate initial and boundary conditions. The only parameter is R, the *Reynolds number*, which is inversely proportional to the viscosity of the fluid. Concerning the derivation of (9.41), we can only remark that when $R\to\infty$ we formally retrieve the inviscid flows described in Chapter 2, and when $R\to 0$ and $p=O(R^{-1})$ we return to the slow flows of Chapter 5. A systematic derivation can be found in [29], where the relevance of the boundary layer equation (6.77) for large finite R is also explained.

In this book we have already seen that, as a result of various approximations, (9.41) is the progenitor of many systems of pdes, some elliptic and some hyperbolic. Nonetheless, there are no rigorous results concerning existence, uniqueness or well-posedness for the full system unless R is quite small. Unfortunately, many interesting and important flows occur at large values of R (say 10^8 in aerodynamics) and, indeed, the phenomenon of turbulence is widely believed to be described by (9.41) with R large. Thus, it is not surprising that (9.41) is so challenging theoretically and we begin our discussion by only looking at the formal limit $R=\infty$.

9.4.4 Inviscid flow: the Euler equations

Adopting our standard philosophy, we begin by looking for real characteristic surfaces for the system

$$\frac{\partial\mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}=-\nabla p,\quad \nabla\cdot\mathbf{u}=0. \tag{9.42}$$

When we 'freeze' the operator $\mathbf{u} \cdot \nabla$ locally in the vicinity of some point of the flow, we find that, when we write $\phi(\mathbf{x}, t)$ = constant as a characteristic surface with $\partial\phi/\partial t = \xi_0$ and $\partial\phi/\partial x_i = \xi_i$, the normal cone is

$$\left(\xi_0 + \sum_{i=1}^{3} u_i \xi_i\right)\left(\sum_{i=1}^{3} \xi_i^2\right) = 0. \tag{9.43}$$

Hence, there is again only one real component of the characteristic surface, and it degenerates into the line

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}, \tag{9.44}$$

which is called the ***particle path***. Moreover, it is easy to see that, in steady flow in which $\partial\mathbf{u}/\partial t$ is zero in (9.42), the quantity $p + \frac{1}{2}|\mathbf{u}|^2$ is a Riemann invariant on these particle paths, which are then called ***streamlines*** (t then simply parametrises distance along the paths). This result is one version of what is called ***Bernoulli's equation*** in hydrodynamics and this simple exercise in Riemann invariants is of fundamental importance in calculating hydrodynamic forces. However, more information can be gleaned from (9.42) by looking at the following generalisation of the idea of Riemann invariants.

In Chapter 4, we simply set ourselves the task of finding functions of the dependent and independent variables that are conserved along characteristics, at least in two-dimensional problems. But why should we not seek functions, even vector or matrix-valued ones, of the ***derivatives*** of $\mathbf{u}$ that might be conserved along characteristics?[181] At first sight this seems a tiresome idea because not only is the evaluation of the equations to be satisfied by the derivatives tedious, but the introduction of ever higher derivatives leads to ever more overdetermined systems for these derivatives. However, look at what happens when we just consider the first spatial derivatives of $\mathbf{u}$ and p in (9.42): if we nimbly exploit the homogeneity of the nonlinear term and eliminate p by cross-differentiation, in two space dimensions (x, y), where $\mathbf{u} = (u, v)$, we obtain

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\left(\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\right) = 0. \tag{9.45}$$

Hence the vorticity

$$\omega = \frac{\partial u}{\partial y} - \frac{\partial v}{\partial x} \tag{9.46}$$

is conserved along the characteristic; when the flow is steady, the Bernoulli function $p + \frac{1}{2}(u^2 + v^2)$ is also conserved along the characteristic.

The generalisation of this idea to three-dimensional flows is even more fascinating. We can either be motivated by (9.46) or prompted by our earlier discussion of hydrodynamics to define

$$\boldsymbol{\omega} = \nabla \wedge \mathbf{u}, \tag{9.47}$$

[181] The idea of looking for conserved derivatives can also be applied to the group theory approach of §6.5; then the groups act on what are called ***jet spaces***.

and now the cross-differentiation exercise yields

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\mathbf{u} \cdot \nabla)\boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \nabla)\mathbf{u} = 0, \tag{9.48}$$

and this equation has implications as dramatic as those of (9.45) for the theory of inviscid hydrodynamics. We have already seen in §9.2.3 that the linearised equation of inviscid flow has *irrotational* solutions, i.e. ones in which $\boldsymbol{\omega} = 0$, if $\boldsymbol{\omega}$ vanishes initially. We can now see that this result also applies to the full nonlinear Euler equations. This is trivial to see from (9.45) and less easy from (9.48); in any case, as anticipated in Chapter 5, there is a large class of irrotational flows for which there exists a velocity potential ϕ such that

$$\mathbf{u} = \nabla\phi. \tag{9.49}$$

For these *potential flows*, (9.42) implies that the nonlinear Euler equations have been reduced to the simplest example of Chapter 5, namely

$$\nabla^2\phi = 0. \tag{9.50}$$

It can also be easily shown, by considering its gradient, that for potential flows the quantity

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + p \tag{9.51}$$

is a global invariant, not just a constant on any one characteristic, as would be obtained by the argument above.

We will say more about the velocity potential later, but let us first return to the rather mysterious equation (9.48), which does not, at first sight, appear to be a conservation statement along the characteristics (9.44). However, we must remember that we are now dealing with the vector $\boldsymbol{\omega}$ and this forces us to reconsider what we mean by conservation along a curve. As illustrated in Fig. 9.3, it only makes geometric sense to say that $\boldsymbol{\omega}$ is conserved along (9.44) if its value at time $t + \delta t$ is related to its value at time t according to that diagram. We begin by drawing the vector $\boldsymbol{\omega}(t)$ through P, consider the evolution of P and the nearby point P' to Q and Q', respectively, and demand that $\boldsymbol{\omega}(t + \delta t)$ is in the direction QQ'. A

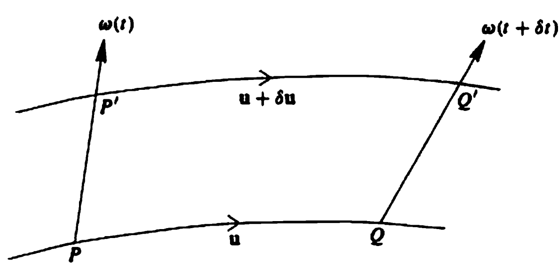


Fig. 9.3 Conservation of vorticity.

simple calculation (Exercise 9.11) shows that (9.48) is precisely the condition for $\boldsymbol{\omega}$ to be conserved in this geometric sense.

We remark that even this idea of conservation needs to be taken one step further when considering fluids such as liquid crystals, whose anisotropic structure requires their properties at any 'particle' to be described by a matrix or linear transformation which is transported by that particle. This then necessitates consideration of conservation of a linear transformation along a vector field such as (9.44). Hence we require a characterisation of matrices $\mathbf{A}$ such that, whenever $\mathbf{a}$ and $\mathbf{a}'$ are conserved, i.e.

$$\frac{\partial \mathbf{a}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{a} = (\mathbf{a}\cdot\nabla)\mathbf{u}, \quad \frac{\partial \mathbf{a}'}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{a}' = (\mathbf{a}'\cdot\nabla)\mathbf{u}, \tag{9.52}$$

and $\mathbf{A}\mathbf{a} = \mathbf{a}'$ at $t = 0$, then $\mathbf{A}\mathbf{a} = \mathbf{a}'$ for all t. As shown in Exercise 9.12, this requires that

$$\frac{\partial \mathbf{A}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{A} = \mathbf{A}\boldsymbol{\Omega} - \boldsymbol{\Omega}\mathbf{A}, \tag{9.53}$$

where the vorticity matrix $\boldsymbol{\Omega}$ has entries

$$\Omega_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right).$$

We conclude our brief discussion of the Euler equations by making some speculations about possible boundary conditions to go with the initial condition in $\mathbf{u}$ that we will clearly need. For irrotational flow we would usually have Neumann data for the elliptic equation (9.50), but in the rotational case the situation is much less clear. We can, however, proceed to make an educated guess on the basis of ideas of transmission of information along characteristics, as well as ideas from ellipticity in Chapter 5. If (9.42) was hyperbolic, we would expect to have to know four pieces of Cauchy data prescribing $\mathbf{u}$ and p at the boundary, or six pieces if we worked with $\mathbf{u}$ and $\boldsymbol{\omega}$. However, from (9.43) we see that there is a single real component of the ray cone, namely the particle path, as well as the complex component associated with the Laplace operator. Moreover, there is a time-like direction, which means that information is only transmitted along a particle path in the direction of the flow. Now we recall that, when $\boldsymbol{\omega} = \mathbf{0}$ on any inlet at which particle paths enter the domain of interest, the problem reduces to Laplace's equation, for which only *one* piece of Cauchy data (typically $\mathbf{u}\cdot\mathbf{n}$) is needed. Hence, we surmise that, when $\boldsymbol{\omega} \neq \mathbf{0}$, we need more information than just the value of $\mathbf{u}\cdot\mathbf{n}$. We cannot give the precise specification here but, as shown in Exercise 9.13, a prescription of $\boldsymbol{\omega}$ itself would lead to an overdetermined problem.

9.4.5 Viscous flow

As hinted earlier, there is little that can be said in this text about the horrifyingly difficult Navier–Stokes system (9.41). As in our earlier example of ohmic heating, we find that, for finite values of R, (9.41) is parabolic and that the only real characteristic manifolds are $t =$ constant. Also (9.48) becomes

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\mathbf{u}\cdot\nabla)\boldsymbol{\omega} - (\boldsymbol{\omega}\cdot\nabla)\mathbf{u} = \frac{1}{R}\nabla^2\boldsymbol{\omega}; \tag{9.54}$$

this suggests that viscosity acts to diffuse the vorticity, which is then no longer conserved in the sense of Fig. 9.3.

Apparently, the only remaining avenues are to look for explicit solutions, should the geometry be symmetric enough, or to seek estimates to help with existence and uniqueness results (see, for example, [14]), or to follow the fertile approach of looking for approximate solutions. Indeed, we have come across several such approximations at various places in this book, but they have all been quite crude compared to the varied and often delicate asymptotic theories that have been developed in recent decades. However, one striking exact result has recently emerged in the special case of two-dimensional steady flow. This is the realisation that is then possible to write down the general solution of the two-dimensional Navier-Stokes equations in terms of two arbitrary functions! The details of this result are too complicated to describe here[182] but they start with the idea of conformal invariance. We recall that in two dimensions the general solution of Laplace's equation can be written as

$$u = \Re f(z),$$

where f is an arbitrary analytic function, while in §5.8.4 we saw that the general solution of the biharmonic equation is

$$u = \Re\left(\bar{z}g(z) + h(z)\right),$$

where g and h are arbitrary analytic functions. Now, in steady flow, (9.54) can be thought of as a nonlinear generalisation of the biharmonic equation because, if we introduce a stream function ψ such that $\mathbf{u} = (\partial\psi/\partial y, -\partial\psi/\partial x)$, then we find

$$\frac{1}{R}\nabla^4\psi = \frac{\partial(\psi, \nabla^2\psi)}{\partial(y,x)}. \tag{9.55}$$

Moreover, we recall from §5.12 that even the nonlinear Liouville equation could be solved by regarding z and $\bar{z}$ as independent variables, leading to its reduction to a Ricatti equation. The same sort of idea works here; (9.55) can be written as

$$\frac{4}{R}\frac{\partial^4\psi}{\partial z^2\partial\bar{z}^2} - 2\mathrm{i}\left(\frac{\partial\psi}{\partial z}\frac{\partial}{\partial\bar{z}} - \frac{\partial\psi}{\partial\bar{z}}\frac{\partial}{\partial z}\right)\frac{\partial^2\psi}{\partial z\partial\bar{z}} = 0,$$

which in turn can be split into the sum of a function and its conjugate as

$$\left\{\frac{2}{R}\frac{\partial^4\psi}{\partial z^2\partial\bar{z}^2} + 2\mathrm{i}\left(\left(\frac{\partial^2\psi}{\partial z\partial\bar{z}}\right)^2 + \frac{\partial\psi}{\partial z}\frac{\partial^3\psi}{\partial z\partial\bar{z}^2}\right)\right\}$$
$$+\left\{\frac{2}{R}\frac{\partial^4\psi}{\partial z^2\partial\bar{z}^2} - 2\mathrm{i}\left(\left(\frac{\partial^2\psi}{\partial z\partial\bar{z}}\right)^2 + \frac{\partial\psi}{\partial\bar{z}}\frac{\partial^3\psi}{\partial z^2\partial\bar{z}}\right)\right\} = 0.$$

[182] See the paper 'Parametrization of general solutions for the Navier-Stokes equations'. *Quart. Appl. Math.*, **52**, 335–41, 1994, by K. B. Ranger.

Each term in curly brackets is a perfect second differential so that

$$\Re\left[\frac{\partial^2}{\partial\bar{z}^2}\left(\frac{2}{R}\frac{\partial^2\psi}{\partial z^2}+\mathrm{i}\left(\frac{\partial\psi}{\partial z}\right)^2\right)\right]=0,$$

and this eventually enables ψ to be written in terms of two arbitrary functions of z and $\bar{z}$.

9.5 Interaction between media

As pdes are used to model more and more complicated configurations in science and technology, we inevitably encounter composite problems where different pdes have to be conjoined in some form or other. A now-classic example, motivated by the study of flutter in aircraft structures or the underwater acoustics of ships and submarines, is the following class of problem.

9.5.1 Fluid/solid acoustic interactions

The simplest situation we might envisage is that of a 'fluid-loaded' elastic solid: the solid itself can transmit waves but so can the fluid in which it is immersed, and a vital question concerns which medium is preferred by sound waves excited locally near the solid. Typically, in two space dimensions in the frequency domain, we may have to solve a Helmholtz equation in an inviscid fluid $y > 0$ with a wave operator in the boundary instead of the conventional Dirichlet or Neumann data discussed in Chapter 5. The simplest configuration is when the solid is a membrane. Then the fluid velocity potential is $\Re\left(e^{-\mathrm{i}\omega t}\phi(x,y)\right)$ and ϕ satisfies

$$(\nabla^2+k^2)\phi=0 \quad \text{for } y>0. \tag{9.56}$$

If the transverse displacement of the membrane is u, the boundary condition is

$$c^2\frac{\partial^2 u}{\partial x^2}+k^2u=\nu p \quad \text{on } y=0; \tag{9.57}$$

here c is the wave speed in the membrane relative to that in the fluid, ν is a positive constant which measures the 'fluid loading', and p is the pressure exerted by the fluid. To close the model we need to use Bernoulli's equation on $y = 0$ and a kinematic condition which, for small disturbances, gives

$$p=\mathrm{i}\omega\phi(x,0), \quad -\mathrm{i}\omega u=\frac{\partial\phi}{\partial y}(x,0), \tag{9.58}$$

together with a specification of the source and any radiation condition that may be necessary.

This type of problem offers many new challenges and opportunities. On the negative side, the radiation condition is even less obvious than it was for the systems considered in Chapter 5, but equally the problem is linear and hence susceptible to Green's function techniques and, in this geometry, to Fourier transforms. As often happens with seemingly simple practical problems like this, the technical details are quite formidable and would require more space than we have here. However, the predictions can be quite surprising and it turns out that, for a source localised in the membrane, it is only when $c < 1$ that this source radiates energy uniformly to infinity in the fluid. This is a simple generalisation of the difference between supersonic and subsonic flow (see Exercise 9.15).

9.5.2 Fluid/fluid gravity wave interaction

Whenever two media interact at a common interface that moves appreciably, a free boundary model can be formulated for the motion of the interface. One common situation where this happens is when immiscible inviscid fluids of different density flow past each other, as discussed in §7.1. Moreover, when the interface motion is so small that the problem can be linearised as in §7.2, all such composite problems can, in principle, be 'condensed' into a statement that involves variables defined only on the boundary between the different regions that are involved. For example, if in (9.56) we write the Dirichlet-to-Neumann operator

$$\frac{\partial \phi}{\partial y}(x,0) = \mathcal{L}\phi(x,0), \tag{9.59}$$

then (9.57) becomes

$$\left(c^2\frac{\mathrm{d}^2}{\mathrm{d}x^2} + k^2\right)\mathcal{L}\phi(x,0) = \nu\omega^2\phi(x,0). \tag{9.60}$$

Of course, $\mathcal{L}$ is now a global operator[183] and we are immediately led into the realm of integro-differential equations, which is, in principle, beyond the scope of this book. However, there are many interesting new ideas involved and the derivation of one spectacular example is described in Exercise 9.16; it is the *Benjamin–Ono* equation,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \mathcal{H}\left(\frac{\partial^2 u}{\partial x^2}\right), \tag{9.61}$$

where $\mathcal{H}$ denotes the Hilbert transform. We will mention it again in §9.7.

9.6 Gauges and invariance

At several stages throughout the book we have encountered situations where pdes have been more or less 'solved' not by integration but by insouciance. This happens, for example, in two-dimensional incompressible fluid dynamics, where we stated boldly that the equation of mass conservation $\partial u/\partial x + \partial v/\partial y = 0$ is automatically 'solved' by the existence of a stream function ψ such that $u = \partial\psi/\partial y$ and $v = -\partial\psi/\partial x$. Most recently, we made the same statement concerning the velocity potential ϕ in (9.49). At a slightly higher level, after (5.110) we asserted that one

[183] In fact,

$$\mathcal{L}u(x) = \int_{-\infty}^{\infty} H_0^{(1)}(k(x-\xi))u(\xi)\,\mathrm{d}\xi,$$

where $H_0^{(1)}$ is the Hankel function of the first kind.

of Maxwell's equations, $\nabla \cdot \mathbf{H} = 0$, could be 'solved' using the existence of a vector potential $\mathbf{A}$ such that $\mathbf{H} = \nabla \wedge \mathbf{A}$ (or of a skew-symmetric matrix whose divergence is $\mathbf{H}$). Quantities like ψ, ϕ and $\mathbf{A}$ are often called *gauges*. Before mentioning some other interesting examples, we can make two obvious general remarks.

1. Most pdes do not possess gauges, or at least not in an obviously identifiable form.
2. There is always a price to pay for using gauges because of their non-uniqueness. In the cases mentioned above, ψ, ϕ and $\mathbf{A}$ are arbitrary to within additive constants and an additive gradient of a function, respectively, while $\mathbf{A}$ could be pinned down to within a constant by requiring that, in the time-independent case, $\nabla \cdot \mathbf{A} = 0$, in which case it is called the *Coulomb gauge*. To ensure well-posedness for models for these gauges in the sense of Chapter 2, it is necessary to relax the requirement of uniqueness to allow for these transformations.

Nevertheless, it is a matter of common experience that the convenience of reducing the size of a system of pdes usually more than outweighs any disadvantages resulting from introducing the gauge and, indeed, gauges may play a vital role in modelling practical problems.

An interesting illustration of the complexity of the idea of gauge functions comes from asking the following question: 'How can we characterise three-component functions $\mathbf{u}(x, y, z)$ in terms of the number of independent scalar functions that are needed to define them?' A more informative answer than simply saying 'three, one for each of the three components of $\mathbf{u}$' is provided by the following hierarchy of vector fields.

1. Suppose the 'vorticity' vanishes:

$$\nabla \wedge \mathbf{u} = \mathbf{0}. \tag{9.62}$$

This is a highly degenerate first-order system of pdes whose normal cone is the whole space; the calculation in §2.6 yields zero for any (ξ_1, ξ_2, ξ_3). Nevertheless, it is well known that a necessary and sufficient condition for (9.62) to hold is

$$\mathbf{u} = \nabla\phi \tag{9.63}$$

for some scalar gauge function ϕ which is only determined to within a constant. Hence $\mathbf{u}$ is characterised analytically by one function, to within a constant, and geometrically by the fact that the field $\mathbf{u}(\mathbf{x})$ is everywhere normal to a family of surfaces $\phi =$ constant. However, this geometric condition would also be satisfied if

$$\mathbf{u} = \phi_1 \nabla \phi_2 \tag{9.64}$$

for some functions ϕ_1 and ϕ_2. This leads us to the next layer in the hierarchy, which follows.

2. Suppose the *helicity*, $\mathbf{u} \cdot \nabla \wedge \mathbf{u}$, vanishes:

$$\mathbf{u} \cdot \nabla \wedge \mathbf{u} = 0. \tag{9.65}$$

This scalar equation is automatically satisfied by (9.64) and is clearly even more underdetermined than (9.62). However, it can be shown that (9.65) implies

(9.64), where ϕ_2 is undetermined, at least to within a multiplicative constant. Hence it follows that (9.65) is necessary and sufficient for the existence of a family of two-dimensional surfaces to which $\mathbf{u}$ is everywhere orthogonal. If (9.65) appeared as part of a system of pdes for $\mathbf{u}$, it could be replaced by the existence of *two* gauge functions ϕ_1 and ϕ_2.

3. The least degenerate situation is when

$$\mathbf{u} \cdot \nabla \wedge \mathbf{u} = f \tag{9.66}$$

is a given function f that does not vanish in the region of interest. Then the best that can be said is that any characterisation of $\mathbf{u}$ requires three scalar functions and, generalising (9.64), these could be the so-called *Clebsch potentials*, which represent $\mathbf{u}$ in the form

$$\mathbf{u} = \phi_1 \nabla \phi_2 + \nabla \phi_3. \tag{9.67}$$

The gauge invariance of ϕ_1, ϕ_2 and ϕ_3 is now even less clear than when the helicity vanishes.

We cannot discuss these questions further here but merely remark that helicity has an important role to play in magneto-hydrodynamics [28]. Even for the Euler equations (9.42) and (9.48) with no boundaries, it is easy to see that, with $\nabla \wedge \mathbf{u} = \boldsymbol{\omega}$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int \mathbf{u} \cdot \boldsymbol{\omega} \, \mathrm{d}\mathbf{x} &= \int \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) (\mathbf{u} \cdot \boldsymbol{\omega}) \, \mathrm{d}\mathbf{x} \\ &= \int (-\nabla p \cdot \boldsymbol{\omega} + \mathbf{u} \cdot (\boldsymbol{\omega} \cdot \nabla) \mathbf{u}) \, \mathrm{d}\mathbf{x} \\ &= \int \nabla \cdot (-p\boldsymbol{\omega} + (\boldsymbol{\omega} \cdot \mathbf{u}) \mathbf{u}) \, \mathrm{d}\mathbf{x}; \end{aligned} \tag{9.68}$$

this vanishes, assuming p and $\mathbf{u}$ decay sufficiently rapidly at infinity, and hence the total helicity is a conserved quantity.

9.7 Solitons

The Benjamin–Ono equation (9.61) is one example of a small class of nonlinear models, most of them pdes, which have had an influence on science out of all proportion to their numbers. One other example that is simple enough to derive here concerns the suspension of a large number of rigid pendulums (e.g. paper clips) from an elastic object with torsional stiffness (e.g. an elastic band) aligned along the x axis, as in Fig. 9.4. In the absence of gravity, torsional waves could propagate down the system according to the well-known one-dimensional wave equation

$$c^2 \frac{\partial^2 \theta}{\partial x^2} - \frac{\partial^2 \theta}{\partial t^2} = 0, \tag{9.69}$$

where $\theta(x, t)$ is the angular displacement of the pendulum at the point x from the downward vertical and c is the torsional wave speed. Equally, in the absence of

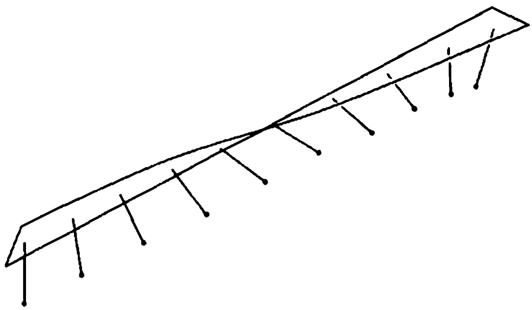


Fig. 9.4 Realisation of the Sine–Gordon equation.

any spatial variation but in the presence of gravity, a pendulum of length l would respond according to the simple pendulum equation

$$\frac{\partial^2 \theta}{\partial t^2} + \frac{g}{l}\sin\theta = 0.$$

Putting these two mechanisms together, we find the famous *Sine–Gordon equation*[184]

$$c^2\frac{\partial^2 \theta}{\partial x^2} - \frac{\partial^2 \theta}{\partial t^2} = \frac{g}{l}\sin\theta. \tag{9.70}$$

Thus far, the only two obvious remarks that this book would have had to say about the Sine-Gordon equation are that it is hyperbolic (we would not even have been able to say that much about the Benjamin–Ono equation) and that there are travelling wave solutions involving hyperbolic secants (and hyperbolic tangents for (9.61)). The discovery that the Cauchy initial value problem can be solved almost explicitly for the Sine–Gordon and Benjamin–Ono equations, and for some other equations to be mentioned later, came about through a coincidence of several seemingly unrelated trains of thought, involving conservation laws, group transformations, eigenvalue problems, scattering theory and inverse problems; the greatest stimulus came from numerical experiments concerning the remarkable way in which certain travelling wave solutions, called *solitons*, can 'pass through' each other and emerge unscathed, much as if they were solutions of the prototype wave equation (9.69). Even thirty years after the pioneering work, it is still more difficult to motivate the key ideas of soliton theory than any of the others in the book, partly because of the absence of any elementary geometrical interpretation.

The closest we have come to soliton theory in the preceding chapters has been in Chapter 6 when we described the possibility of using groups of transformations to simplify certain pdes. There, we only sought simple 'similarity transformations' of the dependent and independent variables. Even the generalisation of the ideas of §6.5 to include derivatives of the dependent variable would soon have led us into

[184]This equation arises in far less transparent situations than this, such as for the variation of the order parameter in a small gap (Josephson junction) between two Type-II superconductors.

unjustifiably complicated technicalities. However, there is no limit to the imagination that might be exercised when asking about objects that might be left invariant when a pde is satisfied. Simple answers to this question that we have seen have been that the object could be a function (e.g. a Riemann invariant) or a functional (such as the integrated density in a conservation law). But suppose we were to ask a question that might be posed by a scientist interested in quantum mechanics: 'could an *operator* that depends on an unknown function be invariant when that function evolves according to some pde?' This is a much more recherché question than its predecessors, as can be seen if we consider the simplest non-trivial linear operators, namely 2×2 matrices. In particular, we ask 'are there any odes that $a_{ij}(t)$ must satisfy so that

$$\mathbf{A}(t) = (a_{ij}(t)) \tag{9.71}$$

evolves in such a way that its eigenvalues and hence its invariants are constant in time?' Not only might this be an interesting question in its own right, but it also leads to the possibility of finding out something about the solution of the differential equations satisfied by a_{ij} by working backwards from the eigenvalues.

Thus, suppose

$$\mathbf{A}\mathbf{x}(t) = \lambda \mathbf{x}(t),$$

where all the eigenvalues λ are independent of t. Then there exists an invertible matrix $\mathbf{C}(t)$ such that

$$\mathbf{A}(t) = \mathbf{C}^{-1}(t)\mathbf{A}(0)\mathbf{C}(t).$$

Hence

$$\begin{aligned}\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{C}^{-1}\right)\mathbf{A}(0)\mathbf{C} + \mathbf{C}^{-1}\mathbf{A}(0)\frac{\mathrm{d}\mathbf{C}}{\mathrm{d}t} \\ &= \mathbf{BA} - \mathbf{AB},\end{aligned} \tag{9.72}$$

where

$$\mathbf{B} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{C}^{-1}\right)\mathbf{C} = -\mathbf{C}^{-1}\frac{\mathrm{d}\mathbf{C}}{\mathrm{d}t},$$

since $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$. Note that

$$(\mathbf{A} - \lambda\mathbf{I})\left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{B}\mathbf{x}\right) = \mathbf{0},$$

so that

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{B}\mathbf{x} = T(t)\mathbf{x} \tag{9.73}$$

for some scalar function $T(t)$. Note also that, if $\mathbf{C}$ were orthogonal,[185] in which case $\mathbf{B}^{\mathsf{T}} = -\mathbf{B}$, then we would find from the identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{x}^{\mathsf{T}}\mathbf{x}\right) = \mathbf{x}^{\mathsf{T}}\left(\mathbf{B}\mathbf{x} + T\mathbf{x}\right) + \left(\mathbf{x}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}} + T\mathbf{x}^{\mathsf{T}}\right)\mathbf{x}$$

[185] If any of the matrices, eigenvalues or eigenvectors is complex, transpose must be replaced by complex conjugate transpose, and C would have to be *unitary*, so that $\overline{\mathbf{C}}^{\mathsf{T}}\mathbf{C} = \mathbf{I}$.

that, as long as $|\mathbf{x}|^2$ is independent of time, $T(t) = 0$ and the evolution of $\mathbf{x}$ would be governed by

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{B}\mathbf{x}. \tag{9.74}$$

The calculation leading to (9.72) is the motivation for the famous *Lax formulation* of soliton theory,[186] which requires the generalisation of (9.72) to general linear operators. To illustrate the finite-dimensional theory, we can construct a system of odes for the a_{ij} in (9.71) by choosing any 2×2 matrix $\mathbf{B}$ whose entries are functions of a_{ij}. But we must be careful about how many of the a_{ij} we allow to vary. If three were constant, from (9.72) we would obtain four odes for the one variable entry; moreover, since the eigenvalues would only depend on this one entry, it would have to be constant. The other extreme would be to allow all four entries to vary and obtain four odes for them, with the constancy of the eigenvalues ensuring the existence of two conserved quantities. It is easiest to look at the intermediate case

$$\mathbf{A} = \begin{pmatrix} a_{11} & x(t) \\ y(t) & a_{22} \end{pmatrix}, \tag{9.75}$$

where a_{11} and a_{22} are given constants. Then it is easy to see that, no matter what choice we make for $\mathbf{B}$, provided only that it is chosen to ensure that $\mathbf{A}$ is of the form (9.75), we end up with $\mathrm{d}x/\mathrm{d}t = F(x, y)$ and $\mathrm{d}y/\mathrm{d}t = -(y/x)F(x, y)$, so that xy, which determines both eigenvalues, is indeed constant.

Now let us extend this argument to the infinite-dimensional case. Suppose $\mathcal{A}$ is a linear differential operator in x with coefficients that involve functions of x and t, and again, to keep the spectrum of $\mathcal{A}$ constant in time, we require that

$$\frac{\partial \mathcal{A}}{\partial t} = \mathcal{B}\mathcal{A} - \mathcal{A}\mathcal{B}. \tag{9.76}$$

This is to be interpreted as saying that

$$\frac{\partial \mathcal{A}}{\partial t}\phi = (\mathcal{B}\mathcal{A} - \mathcal{A}\mathcal{B})\,\phi$$

for arbitrary smooth functions $\phi(x, t)$. Now, however, we demand further that these functions are such that $(\mathcal{B}\mathcal{A} - \mathcal{A}\mathcal{B})\phi$ does not involve any *differentiation* of ϕ but is purely ϕ multiplied by functions involving the coefficients in $\mathcal{A}$ and $\mathcal{B}$; hence we will be led to one or more pdes for the coefficient functions in $\mathcal{A}$ and $\mathcal{B}$. (This corresponds to $\mathbf{BA} - \mathbf{AB}$ being a square matrix in the finite-dimensional case; we could not have carried through example (9.75) unless $\mathbf{B}$ were a 2×2 matrix). In particular, for historical reasons, we could choose

$$\mathcal{A} = -\frac{\partial^2}{\partial x^2} + u(x, t), \quad \text{so that} \quad \frac{\partial \mathcal{A}}{\partial t} = \frac{\partial u}{\partial t},$$

[186]Note the analogy between (9.72) and (9.53) and, when $\mathbf{C}$ is orthogonal, with the idea of angular velocity in classical mechanics.

and

$$\mathcal{B} = -4\frac{\partial^3}{\partial x^3} + 3\left(u\frac{\partial}{\partial x} + \frac{\partial}{\partial x}u\right).$$

We would then find that (9.76) is the famous *Korteweg–de Vries (KdV) equation*[187]

$$\frac{\partial u}{\partial t} + \frac{\partial^3 u}{\partial x^3} = 6u\frac{\partial u}{\partial x} \tag{9.77}$$

and equations such as (9.70) can emerge similarly when more complicated choices are made for $\mathcal{A}$.

The way is now open to construct the following intriguing recipe proposed for the Cauchy problem for (9.77).

1. Given $u(x,0) = u_0(x)$, find the eigenvalues of the operator $\mathcal{A}(u_0)$ (in the KdV case, $\mathcal{A}(u_0) = -\partial^2/\partial x^2 + u_0$) with suitable boundary conditions.
2. Knowing the constancy of these eigenvalues, try to find $\mathcal{A}$ at a later time and from it solve an inverse problem to read off $u(x,t)$.

This plan is much easier to state than it is to implement, because of the difficulty of step 2. Hence we will restrict attention to the simplest case of the Cauchy problem for the KdV equation with u_0 prescribed and positive in $-\infty < x < \infty$ and decaying sufficiently rapidly as $|x| \to \infty$. Then $u(x,t)$ is such that the eigenvalue problem

$$\mathcal{A}\hat{\psi} = -\frac{\partial^2\hat{\psi}}{\partial x^2} + u(x,t)\hat{\psi} = -\lambda\hat{\psi} \quad \text{for } -\infty < x < \infty, \tag{9.78}$$

with $|\hat{\psi}|$ bounded, has a real spectrum independent of t even though $\mathcal{A}$ evolves in time according to (9.76). Also, because of our assumptions about u, we expect the spectrum to contain all negative values $\lambda = -\omega^2$, and possibly some discrete positive values $\lambda = \omega_n^2$ with corresponding eigenfunctions $\hat{\psi}_{(n)}(x,t)$.

Now comes the key observation that this spectrum describes the modes of oscillation of a one-dimensional elastic continuum modelled by the hyperbolic equation

$$\frac{\partial^2\psi}{\partial x^2} - \frac{\partial^2\psi}{\partial \tau^2} - u(x,t)\psi = 0; \tag{9.79}$$

in fact, setting $\lambda = -\omega^2$, ψ is related to $\hat{\psi}$ by

$$\hat{\psi}(x,\omega;t) = \int_{-\infty}^{\infty} \psi(x,\tau;t)\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau,$$

and, by inversion, ψ is a suitably weighted sum of terms $\hat{\psi}\mathrm{e}^{-\mathrm{i}\omega\tau}$ and $\hat{\psi}_{(n)}\mathrm{e}^{\omega_n\tau}$. Note that this is the first time in the book that we have taken a Fourier transform with respect to a time-like variable but, in one space dimension, time and space

[187]This spectacular equation can be derived from classical inviscid hydrodynamics by combining the ideas of shallow water theory and those of modulated travelling waves (see Exercise 9.16).

are mathematically interchangeable. Indeed, in the following discussion we shall at various points be thinking of both x and τ as the 'time-like' variable.

We remember that we can compute the t derivatives of $\hat{\psi}$ and $\hat{\psi}_{(n)}$, and hence of ψ, from the generalisation of (9.73) to the infinite-dimensional case. However, since t only enters (9.79) as a parameter, we will drop the explicit t dependence of u, ψ and $\hat{\psi}$ for the time being.

The nice thing about (9.79) is its physical interpretation, because we can imagine we are, say, a geologist who wishes to identify an inhomogeneity $u(x)$ from a knowledge of the modes of vibration of that medium. One way to proceed is to oscillate the medium at all frequencies and record its response, and this is conveniently done by exciting a pulse at some point $x = X$, where X is sufficiently large and negative to be remote from the inhomogeneity. The situation is illustrated schematically in Fig. 9.5; the pulse propagates in the positive x direction, and the reflected wave r is observed at $x = X$. Since

$$\psi \sim \delta(x - X - \tau) + r(x + \tau + X) \quad \text{as } x \to -\infty, \tag{9.80}$$

the argument of r being chosen because we anticipate no reflection from $x = O(1)$ until $\tau + X = O(1)$, the relevant Cauchy data to model this excitation is

$$\psi = \delta(-\tau) + r(\tau + 2X), \quad \frac{\partial \psi}{\partial x} = \delta'(-\tau) + r'(\tau + 2X) \quad \text{at } x = X. \tag{9.81}$$

Our knowledge of Riemann functions from Chapter 4 puts us in a good position to analyse this problem. The Riemann function for (9.79) is found by setting

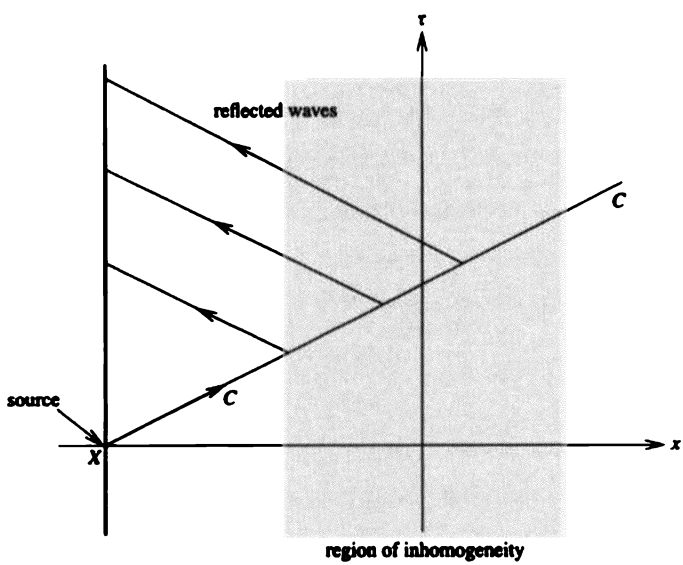


Fig. 9.5 Excitation of the half-space $x > X$; ψ vanishes below the characteristic $x = \tau + X$, denoted by C.

the right-hand side equal to a product of delta functions of x and τ but, from Exercise 4.22, we could equally have considered a suitable combination of delta functions as Cauchy data. For example, if we set ψ_1 to be the solution of (9.79) with

$$\psi_1 = \delta(-\tau), \quad \frac{\partial \psi_1}{\partial x} = \delta'(-\tau) \quad \text{at } x = X, \tag{9.82}$$

we find that $\psi_1 - \delta(x - \tau - X)$ is a bounded function of x, τ and X which, for convenience, we write as

$$\psi_1 - \delta(x - \tau - X) = \Psi_1(x, X, \tau + X), \tag{9.83}$$

where Ψ_1 is non-zero only in $X - x < \tau < x - X$. Now let us write $\xi = x - \tau$ and $\eta = x + \tau$, so that

$$\frac{\partial^2 \psi_1}{\partial \xi \partial \eta} = \frac{1}{4} u \left(\frac{\xi + \eta}{2} \right) \psi_1;$$

since the 'singular support' of ψ_1 is at $x = \tau + X$, we integrate from $\xi = X - 0$ to $\xi = X + 0$ to see that u can be retrieved from a knowledge of Ψ_1 via the equation

$$\lim_{\xi \to X+0} \frac{\partial \Psi_1}{\partial \eta} = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\tau} \Psi_1(\tau + X, X, \tau + X) = \frac{1}{4} u(\tau + X); \tag{9.84}$$

note that Ψ_1 loses its explicit dependence on X as we approach $x = \tau + X$.

Let us return to our geologist, who will find it much harder to simulate (9.82) than (9.81). Let us therefore try to synthesise the solution with data (9.81) from ψ_1 and ψ_2, where

$$\psi_2 = \delta(-\tau), \quad \frac{\partial \psi_2}{\partial x} = -\delta'(-\tau) \quad \text{at } x = X; \tag{9.85}$$

as in (9.83), $\psi_2 = \delta(-x - \tau + X) + \Psi_2(x, X, \tau + X)$, where again Ψ_2 is bounded and non-zero only in $X - x < \tau < x - X$. The data at $x = X$ for the respective Fourier transforms are

$$\begin{aligned} \hat{\psi} &= 1 + \mathrm{e}^{-2\mathrm{i}\omega X} \hat{r}, & \frac{\mathrm{d}\hat{\psi}}{\mathrm{d}x} &= \mathrm{i}\omega(1 - \mathrm{e}^{-2\mathrm{i}\omega X} \hat{r}), \\ \hat{\psi}_1 &= 1, & \frac{\mathrm{d}\hat{\psi}_1}{\mathrm{d}x} &= \mathrm{i}\omega, \\ \hat{\psi}_2 &= 1, & \frac{\mathrm{d}\hat{\psi}_2}{\mathrm{d}x} &= -\mathrm{i}\omega, \end{aligned} \tag{9.86}$$

and so $\hat{\psi} = \hat{\psi}_1 + \hat{r}\hat{\psi}_2 \mathrm{e}^{-2\mathrm{i}\omega X}$. Hence, by the convolution theorem,

$$\begin{aligned} \psi(x, \tau) &= \psi_1(x, \tau) + r(\tau + 2X) * \psi_2(x, \tau) \\ &= \psi_1(x, \tau) + \int_{-\infty}^{\infty} r(\tau' + 2X)\big(\delta(X - x - (\tau - \tau')) \\ &\qquad\qquad + \Psi_2(x, X, \tau - \tau' + X)\big) \, \mathrm{d}\tau' \\ &= \delta(x - X - \tau) + \Psi_1(x, X, \tau + X) + r(\tau + x + X) \\ &\qquad + \int_{\tau - x + X}^{\tau + x - X} r(\tau' + 2X) \Psi_2(x, X, \tau - \tau' + X) \, \mathrm{d}\tau'. \end{aligned} \tag{9.87}$$

We need only make two final observations to turn this into an equation for Ψ_1 in terms of r.

1. The equations for Ψ_1 and Ψ_2 are

$$\frac{\partial^2\Psi_1}{\partial x^2} - \frac{\partial^2\Psi_1}{\partial\tau^2} - u(x)\Psi_1 = \delta(x-\tau-X)u(x),$$
$$\frac{\partial^2\Psi_2}{\partial x^2} - \frac{\partial^2\Psi_2}{\partial\tau^2} - u(x)\Psi_2 = \delta(-x-\tau+X)u(x) = \delta(x+\tau-X)u(x),$$

with zero Cauchy data on $x = X$. Hence

$$\Psi_1(x, X, \tau+X) = \Psi_2(x, X, -\tau+X),$$

so that the integral in (9.87) is

$$\int_{\tau-x+X}^{\tau+x-X} r(\tau'+2X)\Psi_1(x, X, \tau'-\tau+X)\,\mathrm{d}\tau' = \int_{2X-x}^{x} r(\tau''+y)\Psi_2(x, X, \tau'')\,\mathrm{d}\tau'',$$

say, where $y = \tau + X$ and $\tau'' = \tau' - \tau + X$.

2. By causality, $\psi \equiv 0$ in $x - X > \tau$, i.e. $x > y$. Hence, when we denote the asymptotic limit as $X \to -\infty$ of $\Psi_1(x, X, \tau + X)$ by $\widetilde{\Psi}_1(x,\tau)$, we obtain a linear Fredholm integral equation for $\widetilde{\Psi}_1$ in terms of r:

$$\widetilde{\Psi}_1(x,y) + r(x+y) + \int_{-\infty}^{x} r(\tau''+y)\widetilde{\Psi}_1(x,\tau'')\,\mathrm{d}\tau'' = 0. \qquad (9.88)$$

In principle, we can solve this so-called *Marchenko equation* for $\widetilde{\Psi}_1$ in $x > y$ given r, and thus retrieve u from (9.84).

The final piece in the overall jigsaw is a formula for the evolution of r as a function of t. We remember that

$$\psi(x,\tau;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}(x,\omega;t)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\omega + \sum_n \hat{\psi}_{(n)}(x,t)\mathrm{e}^{\omega_n\tau},$$

where, by comparison with (9.80), as $x \to -\infty$,

$$\hat{\psi}(x,\omega;t) \sim \mathrm{e}^{\mathrm{i}\omega(x-X)} + \hat{r}(\omega,t)\mathrm{e}^{-\mathrm{i}\omega(x+X)},$$
$$\hat{\psi}_{(n)}(x,t) \sim \hat{r}_{(n)}(t)\mathrm{e}^{\omega_n x},$$

with $\omega_n > 0$, so that

$$r(x+\tau,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{r}(\omega,t)\mathrm{e}^{-\mathrm{i}\omega(x+\tau)}\,\mathrm{d}\tau + \sum_n \hat{r}_{(n)}(t)\mathrm{e}^{\omega_n(x+\tau)};$$

moreover, $\mathcal{B} \sim -4\partial^3/\partial x^3$ as $x \to -\infty$. To study the evolution in t of $\hat{\psi}$ and $\hat{\psi}_{(n)}$, we must use the infinite-dimensional generalisation of (9.73) or, if we work with eigenfunctions whose norm is independent of t, we can use the simpler equation (9.74). Adopting the former strategy, we find that $\hat{\psi}_{(n)}$ can only evolve as an acceptable eigenfunction if $T(t) = 0$, so that

$$\frac{\partial \hat{\psi}_{(n)}}{\partial t} = \mathcal{B}\hat{\psi}_{(n)}, \quad \text{i.e.} \quad \frac{\mathrm{d}\hat{r}_{(n)}}{\mathrm{d}t} = -4\omega_n^3 \hat{r}_{(n)}, \tag{9.89}$$

while $\hat{\psi}(x,\omega;t)$ must satisfy

$$\frac{\partial \hat{\psi}}{\partial t} = \mathcal{B}\hat{\psi} + T_\omega \hat{\psi},$$

where

$$\mathrm{e}^{-\mathrm{i}\omega(x+X)}\frac{\partial \hat{r}}{\partial t} = -4\mathrm{e}^{-\mathrm{i}\omega X}\left(\mathrm{i}\omega^3 \hat{r}\mathrm{e}^{-\mathrm{i}\omega x} - \mathrm{i}\omega^3 \mathrm{e}^{\mathrm{i}\omega x}\right) + \mathrm{e}^{-\mathrm{i}\omega X} T_\omega \left(\hat{r}\mathrm{e}^{-\mathrm{i}\omega x} + \mathrm{e}^{\mathrm{i}\omega x}\right).$$

Hence,

$$T_\omega = -4\mathrm{i}\omega^3 \quad \text{and} \quad \frac{\partial \hat{r}}{\partial t} = -8\mathrm{i}\omega^3 \hat{r}, \tag{9.90}$$

and $\hat{r}_{(n)}$ and $\hat{r}$ can be found in terms of their initial values.

To recapitulate, the procedure is, given u_0, to

(i) find $\hat{r}_{(n)}(0)$, $\hat{r}(\omega,0)$ and ω_n as a 'direct' problem in spectral theory;

(ii) update $\hat{r}_{(n)}(t)$ and $\hat{r}(\omega,t)$ from (9.89) and (9.90);

(iii) solve (9.88) for $\tilde{\Psi}_1$;

(iv) find $u(x,t)$ from (9.84).

The terms in the solution u that correspond to the discrete spectrum ω_n are called *solitons* (see Exercise 9.17) and they have many fascinating properties [15]. Indeed, they are the simplest of the travelling wave solutions that we mentioned at the beginning of this section and the inverse scattering theory explains beautifully how they eventually emerge from arbitrary initial data and how they interact with each other.

Putting inverse scattering theory into practice is often easier said than done, even in simple cases. The simplest of all is when $u_0 = u(x,0) \equiv 0$. Then $\hat{r}_{(n)}(0) = \hat{r}(\omega,0) = 0$, and so $r = 0$; then (9.88) gives $\tilde{\Psi}_1 = 0$ and (9.85) that $u(x,t) = 0$, with $\psi = \delta(x - X - \tau)$. Less trivially, suppose $u(x,0)$ is so small that r and $\tilde{\Psi}_1$ are also small, and the integral in (9.88) can be neglected. Also suppose there is no discrete spectrum. We first need to find $\hat{r}(\omega,0)$ in terms of u_0, which we do by writing $\hat{\psi}(x,\omega;0) = \mathrm{e}^{\mathrm{i}\omega(x-X)} + \tilde{\psi}$, where

$$\frac{\mathrm{d}^2\tilde{\psi}}{\mathrm{d}x^2} + \omega^2\tilde{\psi} = u_0(x)\mathrm{e}^{\mathrm{i}\omega(x-X)}$$

to lowest order, with $\tilde{\psi} \sim \mathrm{e}^{-\mathrm{i}\omega(x+X)}\hat{r}(\omega,0)$ as $x \to -\infty$ and $\tilde{\psi}$ proportional to $\mathrm{e}^{\mathrm{i}\omega x}$ as $x \to +\infty$. Since, by variation of parameters,

$$2\mathrm{i}\omega\tilde{\psi} = \mathrm{e}^{\mathrm{i}\omega(x-X)}\int_{-\infty}^{x} u_0(\xi)\,\mathrm{d}\xi - \mathrm{e}^{-\mathrm{i}\omega(x+X)}\int_{\infty}^{x} u_0(\xi)\mathrm{e}^{2\mathrm{i}\omega\xi}\,\mathrm{d}\xi,$$

we find

$$\hat{r}(\omega, 0) = \frac{\hat{u}_0(2\omega)}{2i\omega}.$$

Now, from (9.90),

$$r(x+\tau, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-8i\omega^3 t} e^{-i\omega(x+\tau)}\hat{r}(\omega, 0)\,d\omega,$$

and so

$$\Psi_1(\tau + X, X, \tau + X) = -r(2(\tau + X), t) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-8i\omega^3 t} e^{-2i\omega(X+\tau)}\frac{\hat{u}_0(\omega)}{2i\omega}\,d\omega.$$

Finally, reinstating the dependence of u on t,

$$u(\tau + X, t) = 2\frac{d}{d\tau}\Psi_1(\tau + X, X, \tau + X),$$

and so, setting $2\omega = k$,

$$u(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ik^3 t - ikx}\hat{u}_0(k)\,dk,$$

which is just the result of taking a Fourier transform in x of $\partial u/\partial t + \partial^3 u/\partial x^3 = 0$.

This interpretation of inverse scattering theory as a generalisation of Fourier transforms makes such a fitting end piece to our book because it illustrates how a little basic knowledge about pdes can lead to some of the most ingenious mathematical procedures ever devised for their solution.

Exercises

9.1. Suppose $\nabla^2 u(r, \theta) = 0$ in $4 < r^2 < 9$ and we wish to find $u(2, \theta)$ from a knowledge of u and $\partial u/\partial r$ on $r = 3$. Show that, if

$$u(3, \theta) = \frac{3\cos\theta - 1}{10 - 6\cos\theta}, \quad \frac{\partial u}{\partial r}(3, \theta) = \frac{6 - 10\cos\theta}{(10 - 6\cos\theta)^2},$$

then $u(2, \theta) = (2\cos\theta - 1)/(5 - 4\cos\theta)$.
Explain why we are lucky to have been able to do this and why we would have been unable to find $u(1/2, \theta)$ when $\nabla^2 u = 0$ in $1/4 < r^2 < 9$.

Remark. If u and $\partial u/\partial r$ were constant on $r = 3$, we could find $u(R, \theta)$ for any $R > 0$. This suggests we can 'regularise' the inverse problem by taking averages of the data over $r = 3$, which corresponds to neglecting the high-frequency Fourier components in $u(3, \theta)$ and $\partial u/\partial r\,(3, \theta)$, thereby extending the applicability of the Cauchy–Kowalevski theorem.

9.2. Consider a Hele-Shaw free boundary problem of the type described in Chapter 7 (in three dimensions, a porous medium flow without gravity), in which, for $t > 0$, fluid is injected at the origin $\mathbf{x} = \mathbf{0}$ at a constant rate Q into an initial domain $\Omega(0)$ containing the origin.

(i) Setting the pressure $p(\mathbf{x},t)=0$ outside the fluid domain $\Omega(t)$, define

$$u(\mathbf{x},t)=\int_0^t p(\mathbf{x},\tau)\,\mathrm{d}\tau,$$

and show that, as in §7.4.1,

$$\nabla^2 u=\begin{cases}0, & \text{in } \Omega(0)\backslash\{\mathbf{0}\},\\ 1, & \text{in } \Omega(t)\backslash\overline{\Omega}(0),\\ 0, & \text{outside } \overline{\Omega}(t).\end{cases}$$

Show also that u and ∇u are continuous on $\partial\Omega(0)$ and $\partial\Omega(t)$.

(ii) Now fix $t>0$ and set $\phi(\mathbf{x},t)=u(\mathbf{x},t)+(Qt/2\pi)\log|\mathbf{x}|$ in two dimensions ($\phi(\mathbf{x},t)=u(\mathbf{x},t)-Qt/4\pi|\mathbf{x}|$ in three dimensions) to show that ϕ is a constant multiple of the gravitational potential due to mass of constant density occupying the annular region between $\partial\Omega(t)$ and $\partial\Omega(0)$. Noting that $u(\mathbf{x},t)\equiv 0$ outside $\Omega(t)$, deduce that knowledge of the gravitational potential outside a domain, together with the assumption of constant mass density, are not sufficient to determine the domain uniquely.

(iii) In two dimensions, use the Green's function representation for $\phi(z,t)$ ($z=x+\mathrm{i}y$ and $\zeta=\xi+\mathrm{i}\eta$) to show that, outside $\Omega(t)$,

$$\frac{\partial\phi}{\partial x}-\mathrm{i}\frac{\partial\phi}{\partial y}=\frac{1}{2\pi}\iint_{\Omega(t)\backslash\Omega(0)}\frac{\mathrm{d}\xi\,\mathrm{d}\eta}{z-\zeta}=\frac{1}{2\pi}\sum_{n=0}^{\infty}z^{-(n+1)}\iint_{\Omega(t)\backslash\Omega(0)}\zeta^n\,\mathrm{d}\xi\,\mathrm{d}\eta$$

for large $|z|$. Deduce from part (ii) that knowledge of all the moments of a domain need not determine the domain uniquely.

9.3. One way to define a three-dimensional vector distribution $\mathbf{v}$ is by the formula

$$(\mathbf{v},\boldsymbol{\psi})=\int_{\mathbb{R}^3}\mathbf{v}\cdot\boldsymbol{\psi}\,\mathrm{d}\mathbf{x},$$

where $\boldsymbol{\psi}$ is a smooth vector test function, all of whose components vanish rapidly at infinity. Use the formulæ

$$\nabla\cdot(\psi\mathbf{a})=\nabla\psi\cdot\mathbf{a}+\psi\nabla\cdot\mathbf{a},\quad \nabla\cdot(\mathbf{a}\wedge\mathbf{b})=\mathbf{b}\cdot(\nabla\wedge\mathbf{a})-\mathbf{a}\cdot(\nabla\wedge\mathbf{b}),$$

where ψ and $\mathbf{a}$, $\mathbf{b}$ are scalar and vector functions, respectively, to motivate the three definitions

$$(\nabla v,\boldsymbol{\psi})=-(v,\nabla\cdot\boldsymbol{\psi})$$

(in which v is a scalar distribution and the inner product on the right is the usual scalar one),

$$(\nabla\cdot\mathbf{v},\psi)=-(\mathbf{v},\nabla\psi)$$

(in which ψ is a scalar test function), and

$$(\nabla\wedge\mathbf{v},\boldsymbol{\psi})=(\mathbf{v},\nabla\wedge\boldsymbol{\psi}).$$

Show that $\nabla\wedge(\nabla v)\equiv\mathbf{0}$ and $\nabla\cdot(\nabla\wedge\mathbf{v})\equiv 0$ for all scalar and vector distributions v and $\mathbf{v}$, respectively.

9.4. By twice taking the curl of (9.16) with $\mathbf{f} = \mathbf{0}$, show that each component u_i of $\mathbf{u}$ satisfies $\nabla^4 u_i = 0$.

9.5. (i) Using the result from Chapter 5 that

$$\nabla^2\left(\frac{1}{r}\right) = -4\pi\delta(x)\delta(y)\delta(z),$$

show that

$$\iiint_{\mathbb{R}^3} e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d\mathbf{r}}{r} = \widehat{\left(\frac{1}{r}\right)} = \frac{4\pi}{k^2},$$

where $k^2 = k_1^2 + k_2^2 + k_3^2 = |\mathbf{k}|^2$. Note that this result involves the interpretation of divergent integrals if it is to be obtained directly. However, following Exercise 5.32, we can write

$$\begin{aligned}\iiint_{\mathbb{R}^3} e^{-i\mathbf{k}\cdot\mathbf{r}}\frac{d\mathbf{k}}{k^2} &= \pi\iint_{\mathbb{R}^2} e^{-i(k_2y+k_3z)-x\sqrt{k_2^2+k_3^2}}\frac{dk_2\,dk_3}{\sqrt{k_2^2+k_3^2}}\\ &= \pi\int_0^{2\pi}\int_0^{\infty} e^{-\rho x - i\rho(y\sin\phi + z\cos\phi)}\,d\rho\,d\phi\\ &= \frac{2\pi^2}{r}.\end{aligned}$$

(ii) Show that

$$\widehat{\left(\frac{x^2}{r^3}\right)} = \frac{4\pi(k_2^2+k_3^2-k_1^2)}{k^4} = \widehat{\left(\frac{1}{r}\right)} - \frac{8\pi k_1^2}{k^4}$$

and

$$\widehat{\left(\frac{xy}{r^3}\right)} = -\frac{8\pi k_1 k_2}{k^4}.$$

(iii) Suppose that $\mathbf{G}$ is a Green's matrix satisfying

$$\mu\nabla^2\mathbf{G} + (\lambda+\mu)\left(\nabla\left(\frac{\partial G_{i1}}{\partial x_i}\right), \nabla\left(\frac{\partial G_{i2}}{\partial x_i}\right), \nabla\left(\frac{\partial G_{i3}}{\partial x_i}\right)\right) = -\delta(\mathbf{x}-\boldsymbol{\xi})\mathbf{I},$$

where $\mathbf{I}$ is the identity and we are using the summation convention so that

$$\frac{\partial G_{i1}}{\partial x_i} = \frac{\partial G_{11}}{\partial x_1} + \frac{\partial G_{21}}{\partial x_2} + \frac{\partial G_{31}}{\partial x_3},$$

etc., and $x_1 = x$, etc. Show that $\hat{\mathbf{G}}$ is given by (9.21) and, using (i) and (ii), that $\mathbf{G}$ is given by (9.22).

9.6. Show that, in two dimensions, the Green's matrix for (9.16), which is such that $\mathbf{G} = O(\log|\mathbf{x}|)$ as $|\mathbf{x}| \to \infty$, is given by

$$\mathbf{G}(\mathbf{x}) = \frac{1}{2\pi\mu(1+\kappa)}\begin{pmatrix} -\kappa\log|\mathbf{x}| + x_1^2/|\mathbf{x}|^2 & x_1x_2/|\mathbf{x}|^2 \\ x_1x_2/|\mathbf{x}|^2 & -\kappa\log|\mathbf{x}| + x_2^2/|\mathbf{x}|^2\end{pmatrix},$$

where $\kappa = (\lambda + 3\mu)/(\lambda + \mu)$.

9.7. Suppose that $\phi(\xi, y)$ satisfies

$$\frac{\partial^2\phi}{\partial\xi^2} + \frac{\partial^2\phi}{\partial y^2} = 0$$

for $y > 0$, with Cauchy data

$$\phi = \frac{\epsilon}{\xi^2 + \delta^2}, \quad \frac{\partial\phi}{\partial y} = 0$$

on $y = 0$ (compare this with the example on p. 46). Show that, when ξ is 'complexified' as on p. 405, the resulting problem for ϕ in terms of y and η is

$$\frac{\partial^2\phi}{\partial y^2} - \frac{\partial^2\phi}{\partial\eta^2} = 0$$

for $y > 0$, with

$$\phi = \frac{\epsilon}{\delta^2 - \eta^2}, \quad \frac{\partial\phi}{\partial y} = 0$$

on $y = 0$. Draw the characteristic diagram, describe the propagation of the singularities at $\eta = \pm\delta$, $y = 0$, and deduce that the original problem is ill-posed.

9.8. Consider the steady-state thermistor equations of §9.4.1 in the rectangle $0 < x < a$, $0 < y < b$. Suppose that the boundary conditions are

$$\frac{\partial\phi}{\partial n} = \frac{\partial T}{\partial n} = 0 \quad \text{on the sides } y = 0, b,$$

so that these sides are both thermally and electrically insulated, and

$$\phi = 0, \quad T = T_0 \quad \text{at } x = 0, \qquad \phi = V, \quad T = T_0 \quad \text{at } x = a,$$

so that these ends are held at constant temperature T_0 while a potential difference V is applied across the device. Show that there is a one-dimensional symmetric solution in which $\phi = V/2$ on the centre line $x = a/2$, which is also the hottest part of the device, provided that

$$V^2 < 8\int_{T_0}^{\infty} \frac{dt}{\sigma(t)}.$$

Remark. This solution can be shown to be unique. Now suppose the time-independent thermistor equations are solved in a two-dimensional region D whose boundary is divided into four parts on which the boundary conditions above apply, alternately constant Dirichlet and homogeneous Neumann. Because the equations with these special boundary conditions are invariant under conformal maps, D can be mapped onto a rectangle in which the solution is as above. Thus the level curves of T and ϕ coincide, so that $T = T(\phi)$, and the above restriction on V is necessary for existence of a solution, independently of the geometry.

9.9. (i) Show that the time-independent space charge equations of §9.4.2 have a solution in which $\phi = \log r + \theta^2$ and $\rho = -2/r^2$, and that the stream function is $\psi = 2\theta/r^2$.

(ii) Show that ρ can be eliminated from the time-independent space charge equations to give the third-order equation

$$\nabla \cdot \left(\nabla^2 \phi \nabla \phi\right) = 0.$$

Suppose that, in two dimensions, the equipotentials $\phi =$ constant coincide with the level curves of a harmonic function Φ, so that $\phi = F(\Phi)$. Deduce that

$$\frac{\nabla\Phi \cdot \nabla|\nabla\Phi|^2}{|\nabla\Phi|^4} = -\frac{\mathrm{d}}{\mathrm{d}\Phi}\log(F'(\Phi)F''(\Phi)) = G(\Phi), \quad \text{say.}$$

Writing $\Phi = \frac{1}{2}\left(w(z) + \overline{w(z)}\right)$, so that $\nabla\Phi = \frac{1}{2}\left(w' + \overline{w'}, \mathrm{i}(w' - \overline{w'})\right)^{\mathsf{T}}$, show that

$$\frac{w''(z)}{(w'(z))^2} + \frac{\overline{w''(z)}}{\left(\overline{w'(z)}\right)^2} = G(\Phi).$$

Differentiate with respect to z and then $\bar{z}$ to show that the only possible form for G is $G(\Phi) = a\Phi + b$ for real constants a and b. Hence show that

$$\frac{w''(z)}{(w'(z))^2} = \frac{1}{2}aw + k,$$

where k is a complex constant, and show that $b = k + \bar{k}$. Finally, show that w is determined by

$$z = c_1 \int^{w} \mathrm{e}^{-ks - \frac{1}{4}as^2}\,\mathrm{d}s,$$

where c_1 is a constant, and that ϕ and Φ are related by

$$\phi = c_2 \int^{\Phi} \left(\int^{\xi} \mathrm{e}^{-bt - \frac{1}{2}at^2}\,\mathrm{d}t\right)^{1/2} \mathrm{d}\xi,$$

where c_2 is purely real if the inner integral is positive.

Remark. Part (ii) gives a three-parameter family of solutions, parametrised by a, b and $\Im k$ (the constants c_1 and c_2 can be scaled out). Further properties of these solutions are described in the paper 'Congruent harmonic functions and space charge electrostatic fields', *IMA J. Appl. Math.*, **39**, 189–214, 1987, by S. A. Smith. Unfortunately, they do not satisfy any physically convenient boundary conditions.

9.10. (i) Show that, when $\partial\mathbf{u}/\partial t = \mathbf{0}$, (9.42) can be written as

$$\frac{1}{2}\nabla|\mathbf{u}|^2 - \mathbf{u} \wedge (\nabla \wedge \mathbf{u}) = -\nabla p$$

and hence that $p + \frac{1}{2}|\mathbf{u}|^2$ is constant on a characteristic $\mathrm{d}\mathbf{x}/\mathrm{d}t = \mathbf{u}$.

(ii) Show that, if $\mathbf{u} = \nabla\phi$ in (9.42), then the gradient of (9.51) is zero.

9.11. In Fig. 9.3, denote P by $\mathbf{x}(t)$ and P' by $\mathbf{x}(t) + \epsilon\boldsymbol{\omega}(\mathbf{x}(t), t)$. In a small time δt, P moves to Q and P' to Q', where Q is $\mathbf{x}(t) + \mathbf{u}(\mathbf{x}, t)\delta t$ and Q' is $\mathbf{x}(t) + \epsilon\boldsymbol{\omega}(\mathbf{x}(t), t) + \mathbf{u}(\mathbf{x}(t) + \epsilon\boldsymbol{\omega}(\mathbf{x}(t), t))\delta t$. Deduce that

$$QQ' - PP' = \epsilon\boldsymbol{\omega} \cdot \nabla\mathbf{u}\,\delta t,$$

to lowest order in δt. Now show that (9.48) implies that

$$\epsilon\boldsymbol{\omega}(\mathbf{x}(t + \delta t), t + \delta t) - \epsilon\boldsymbol{\omega}(\mathbf{x}(t), t) = \epsilon(\boldsymbol{\omega} \cdot \nabla)\mathbf{u}\,\delta t;$$

deduce that $P'Q' = \epsilon\boldsymbol{\omega}(\mathbf{x}(t + \delta t), t + \delta t)$ and hence that, if PQ lies along $\boldsymbol{\omega}$ at t, then $P'Q'$ lies along $\boldsymbol{\omega}$ at $t + \delta t$.

9.12. Show that, if $\mathbf{a}$ and $\mathbf{a}'$ evolve according to (9.52) and

$$\mathbf{A}\mathbf{a} = \mathbf{a}',$$

then

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\mathbf{a} = ((\mathbf{A}\mathbf{a}) \cdot \nabla)\mathbf{u} - \mathbf{A}(\mathbf{a} \cdot \nabla)\mathbf{u}, \quad \frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + (\mathbf{u} \cdot \nabla).$$

Deduce that, in two dimensions at least, $\mathrm{d}\mathbf{A}/\mathrm{d}t$ satisfies (9.53).

9.13. Suppose that $\mathbf{u}$ and $\boldsymbol{\omega}$ satisfy the degenerate quasilinear system

$$(\mathbf{u} \cdot \nabla)\boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \nabla)\mathbf{u} = \mathbf{0}, \quad \boldsymbol{\omega} = \nabla \wedge \mathbf{u}, \quad \nabla \cdot \mathbf{u} = 0,$$

and that data is to be prescribed on $z = 0$ for a solution in $z > 0$. Show that a knowledge of the tangential components of $\mathbf{u}$ there determines the normal component of $\boldsymbol{\omega}$, and hence that prescribing $\mathbf{u}$ and $\boldsymbol{\omega}$ on $z = 0$ would lead to an overdetermined problem.

Remark. It can be shown that it is sufficient for just three pieces of scalar information to be prescribed on $z = 0$. For example, we could prescribe $\mathbf{u}$ or, more likely, the normal component of $\mathbf{u}$ and the tangential components of $\boldsymbol{\omega}$.

9.14. Suppose that a sprung piston at $x = X(t)$ oscillates with small amplitude about $x = 0$ in a tube under the action of a pressure $-p$. The model is

$$\frac{\mathrm{d}^2 X}{\mathrm{d}t^2} + \Omega^2 X = \nu p,$$

where p is given by the acoustic model

$$p = \left.\frac{\partial\phi}{\partial t}\right|_{x=0}, \qquad \dot{X} = \left.\frac{\partial\phi}{\partial x}\right|_{x=0}, \qquad \frac{\partial^2\phi}{\partial t^2} = a_0^2\frac{\partial^2\phi}{\partial x^2} \quad \text{for } x > 0.$$

Show that solutions are possible in which $\phi = \Re A e^{\mathrm{i}(kx - \omega t)}$, which corresponds to outgoing waves as $x \to +\infty$, as long as $a_0 = \omega/k$ and $\omega = \Omega/\sqrt{1 + \mathrm{i}\nu/k}$. The fact that ω is complex means that the energy radiating to $x = +\infty$ damps the piston motion.

9.15. Suppose ϕ, p and u satisfy (9.56)–(9.58), with $k > 0$. Show that solutions are possible in which

$$\phi = \Re\left(Ae^{-y\sqrt{K^2-k^2}+iKx}\right)$$

for some constant K. It can be shown that these solutions are physically acceptable as long as $\Re\sqrt{K^2-k^2} > 0$ for $|K| > k$, and $\Im\sqrt{K^2-k^2} < 0$ for $|K| < k$, the latter being a manifestation of the radiation condition. Show that

$$\left(cK^2 - k^2\right)\sqrt{K^2-k^2} - \nu k^2 = 0,$$

and hence that, for small ν (i.e. when the membrane is only slightly affected by the fluid), either

$$c < 1 \quad \text{and} \quad K \text{ is close to } \quad \frac{k}{\sqrt{c}}\left(1 + \frac{\nu c}{2k\sqrt{1-c^2}}\right),$$

or $c > 1$ and either

$$K \text{ is close to } \quad \frac{k}{\sqrt{c}}\left(1 + \frac{i\nu c}{2k\sqrt{c^2-1}}\right)$$

or

$$K \text{ is close to } \quad k\left(1 + \frac{\nu^2}{2k^2(c-1)^2}\right).$$

Remark. The physical interpretation of this result is that either waves can propagate 'subsonically' along the membrane and decay as y increases in the fluid, or they can propagate supersonically and either (i) radiate into the fluid and decay as x increases or (ii) not radiate and decay as y increases. (The authors are grateful to Dr R. H. Tew for this example.)

9.16. (i) When water flows in a shallow, nearly horizontal, layer of depth $\eta(x,t)$ with velocity $u(x,t)$, (7.28) and (7.29) become, after expanding ϕ as a power series in y,

$$\frac{\partial \eta}{\partial x} + \frac{\partial u}{\partial t} + \epsilon u \frac{\partial u}{\partial x} = 0,$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial u}{\partial x} + \epsilon\left(\frac{\partial}{\partial x}(u\eta) + \frac{1}{3}\frac{\partial^3 u}{\partial x^3}\right) = 0,$$

respectively, where ϵ is a small parameter. Show that, when we set $x - t = \xi$ and $\tau = \epsilon t$, we obtain

$$\frac{\partial}{\partial \xi}(\eta - u) + \epsilon\left(\frac{\partial u}{\partial \tau} + u\frac{\partial u}{\partial \xi}\right) = 0,$$

$$\frac{\partial}{\partial \xi}(u - \eta) + \epsilon\left(\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial \xi}(\eta u) + \frac{1}{3}\frac{\partial^3 u}{\partial \xi^3}\right) = 0,$$

and deduce the KdV equation in the form

$$2\frac{\partial \eta}{\partial t} + 3\eta\frac{\partial \eta}{\partial \xi} + \frac{1}{3}\frac{\partial^3 \eta}{\partial \xi^3} = 0$$

as the lowest-order approximation for small ϵ.

(ii) When the shallow layer in (i) is overlain by a half-space of light fluid, the equations for u and η can be rescaled to become

$$\frac{\partial \eta}{\partial x} + \frac{\partial u}{\partial t} + \epsilon u \frac{\partial u}{\partial x} = 0,$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial u}{\partial x} + \epsilon \left(\frac{\partial}{\partial x}(\eta u) + \frac{\partial p}{\partial x} \right) = 0,$$

where p is the pressure at the base of the lighter fluid. If the lighter fluid flows irrotationally with velocity potential $\bar{\phi}$, then, after linearisation in the usual way, it follows that, to lowest order,

$$p + \frac{\partial \bar{\phi}}{\partial t} + \eta = 0, \quad \frac{\partial \bar{\phi}}{\partial y} = \frac{\partial \eta}{\partial t} \quad \text{on } y = 0,$$

where $\nabla^2 \bar{\phi} = 0$ in $y > 0$. Show that

$$\frac{\partial \bar{\phi}}{\partial x} = \mathcal{H}\left(\frac{\partial \bar{\phi}}{\partial y} \right) \quad \text{on } y = 0,$$

where $\mathcal{H}$ denotes the Hilbert transform, and follow the argument of (i) to derive the Benjamin–Ono equation in the form

$$2\frac{\partial \eta}{\partial \tau} + 3\eta \frac{\partial \eta}{\partial \xi} - \frac{\partial \eta}{\partial \xi} - \mathcal{H}\left(\frac{\partial^2 \eta}{\partial \xi^2} \right) = 0.$$

9.17. For the KdV equation in the form

$$\frac{\partial u}{\partial t} + 6u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0,$$

show that a travelling wave solution in which $u = u(x - ct)$ exists for all $c > 0$, and that

$$u = \frac{c}{2}\,\mathrm{sech}^2\left(\frac{\sqrt{c}(x - ct)}{2} \right).$$

Show also that $\mathrm{sech}\,(\sqrt{c}(x - ct)/2)$ is the principal eigenfunction for (9.78) in this case.

Remark. It can be shown that, when $u(x,0) = (c/2)\,\mathrm{sech}^2\,(\sqrt{c}x/2)$, the discrete spectrum of (9.78) is the single point $c/4$.

9.18. For the Sine–Gordon equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = \sin u,$$

show that there are travelling wave solutions symmetric about $x = ct$ and that

$$u = \begin{cases} 4\tan^{-1}\tanh\left((x - ct)/2\sqrt{c^2 - 1}\right), & c > 1, \\ 4\tan^{-1}\exp\left((x - ct)/2\sqrt{c^2 - 1}\right), & c < 1. \end{cases}$$

Conclusion

We will conclude by recapitulating what we think are the principal lessons to be learned from an overview of what we have tried to describe in this book.

The first is the concept of well-posedness, so dramatically illustrated by the contrast between the Cauchy problems for elliptic and hyperbolic equations. Even the simplest examples in §3.1 give a reliable guide of how a change in sign in one term in the equation can change all the rules of the game. It is our biggest source of regret that it is only at all easy to give a general answer to the question of well-posedness for hyperbolic equations and that the reality or otherwise of the characteristics can lead to such a dichotomy; maybe in years to come, the theory of complex characteristics will have developed to an extent that this distinction is at least more blurred.

The second, closely related topic is that of the qualitative nature of the solution. It is always important to ask questions like 'can the equation admit "wave solutions" that propagate in the interior of the domain, as in Chapter 4, or does the equation immediately smooth away any irregularities in the data, as in Chapters 5 and 6?'

Concerning the representation of solutions, the only all-embracing concept to emerge is that of the formal solution of an arbitrary linear equation in terms of the inverse of the differential operator. This is an integral of the relevant data weighted with a suitable Green's or Riemann function, unless the Fredholm Alternative dictates otherwise.

No such general principles apply when it comes to writing down explicit solutions of partial differential equations. It is always a red-letter day when an explicit solution emerges in an uncontrived situation and it reflects some kind of symmetry or invariance property. However, the latter may be even harder to discern than the solution itself. In this respect, we can only hope that the reader has not been daunted by the plethora of tricks that have been invented to treat special types of equation on a case-by-case basis, even though the list of those we have described is far from exhaustive; we have only tried to describe those devices that seem to offer most insight and generality.

There are two important aspects of the mathematics of partial differential equations that have been woefully underemphasised in this book. The first is any discussion of 'perturbation methods'. As mentioned in the Introduction, this would have been possible at the cost of a change in character and size. The great potential of the ideas that we have excluded can be glimpsed in Chapter 7, where stability theory was seen to be the basic tool for building up a theoretical framework for free boundary problems, and in Chapter 8, where the Geometric Theory of Diffraction is a vital aid to the understanding of high-frequency wave propagation.

The second lacuna is the lack of any treatment of the relationship between what we have expounded and the understanding that can be obtained from numerical computations. Again, it would have been possible, with help, to have doubled the length of the book by including in each chapter comments about algorithms, convergence, error estimates and stability analysis. Or we could have simply included figures obtained by the easy route of attacking the differential equation with the best available software. Apart from our desire for brevity, our only justification for these omissions is that the delicacy of so many of the situations we have encountered demands the need for a quality control that only a mathematical approach can provide.

We also thought about including more discussion of the implications of the mathematics that motivated the individual equations. In many cases these can be quite startling, ranging from the prediction of unexpected instabilities to suggestions for easy and reliable algorithms that enable processes to be optimised. But this, along with greater emphasis on the theoretical aspects, would have made greater demands on the reader than we wanted to impose.

At the end of the day, the feature of writing this book that has given the authors most pleasure has been the exposition, in fairly simple mathematical terms, of the quantitative understanding of so many phenomena of practical importance in everyday life. We know of no other branch of theoretical science where so many situations could be modelled and analysed in anything like four hundred pages, starting more or less from scratch. However, the seemingly comprehensive success of partial differential equations in this respect must not let students be lulled into a false sense of security. For one thing, they are rather special models that are much better behaved in general terms than, say, discrete models. Also, it must always be remembered how dangerous it is to step away from those situations where analytic equations are used in conjunction with analytic solutions. Although we have only encountered one example (on p. 67) where the whole theory crashes in the face of non-analyticity, the literature contains many cases where artificially introduced singularities can destroy all intuition about classification and qualitative behaviour. Our discussion in Chapter 7 has shown that even the most prosaic examples of models with discontinuous coefficients can lead to situations where, with our current knowledge, there is no systematic way to make mathematical sense of the partial differential equations.

References

The number(s) in parentheses indicate the page(s) on which the reference is cited.

[1] Arnol'd, V. I. (1984). *Catastrophe theory.* Springer-Verlag, New York. [369]

[2] Arscott, F. M. (1964). *Periodic differential equations. An introduction to Mathieu, Lamé, and allied functions.* Pergamon Press, Oxford. [119]

[3] Bender, C. M. and Orszag, S. A. (1978). *Advanced mathematical methods for scientists and engineers.* McGraw-Hill, Tokyo. [88]

[4] Birkhoff, G. and Zarantonello, E. H. (1957). *Jets, wakes and cavities.* Academic Press, New York. [342]

[5] Bleistein, N. (1984). *Mathematical methods for wave phenomena.* Academic Press, London. [364, 373]

[6] Bonnet, M. (1997). *Boundary integral equations methods applied to solid and fluid mechanics.* Wiley, New York. [174]

[7] Buckmaster, J. D. and Ludford, G. S. S. (1982). *Theory of laminar flames.* Cambridge University Press. [316]

[8] Carrier, G. F., Krook, M. and Pearson, C. E. (1966). *Functions of a complex variable. Theory and technique.* McGraw-Hill, New York. [109, 181, 182]

[9] Carslaw, H. S. and Jaeger, J. C. (1959). *Conduction of heat in solids* (2nd edn). Clarendon Press, Oxford. [254]

[10] Chester, C. R. (1971). *Techniques in partial differential equations.* McGraw-Hill, New York. [394]

[11] Coddington, E. A. and Levinson, N. (1955). *Theory of ordinary differential equations.* McGraw-Hill, New York. [396]

[12] Courant, R. and Hilbert, D. (1962). *Methods of mathematical physics,* Vols I & II. Interscience, New York. [47, 169, 184, 334, 383]

[13] Crank, J. (1984). *Free and moving boundary problems.* Clarendon Press, Oxford. [335]

[14] Doering, C. R. and Gibbon, J. D. (1995). *Applied analysis of the Navier–Stokes equations.* Cambridge University Press, New York. [413]

[15] Drazin, P. G. and Johnson, R. S. (1989). *Solitons: an introduction.* Cambridge University Press. [141, 425]

[16] Driscoll, T. A. and Trefethen, L. N. (2002). *Schwarz-Christoffel mapping.* Cambridge University Press. [342]

[17] Elliott, C. M. and Ockendon, J. R. (1982). *Weak and variational methods for moving boundary problems.* Pitman, London. [213, 332, 335, 336]

[18] Engl, H. W., Hanke, M. and Neubauer, A. (1996). *Regularization of inverse problems.* Kluwer, Dordrecht. [393]

[19] Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G. (1954). *Tables of integral transforms*, Vol. I. McGraw-Hill, New York. [124]

[20] Gakhov, F. D. (1966). *Boundary value problems.* Pergamon Press, Oxford. [211]

[21] Garabedian, P. R. (1964). *Partial differential equations.* Wiley, New York. [184, 405]

[22] Hinch, E. J. (1991). *Perturbation methods.* Cambridge University Press. [88, 318, 406]

[23] John, F. (1971). *Partial differential equations.* Springer-Verlag, New York. [68]

[24] Jones, D. S. (1986). *Acoustic and electromagnetic waves.* Clarendon Press, Oxford. [209]

[25] Kevorkian, J. and Cole, J. D. (1981). *Perturbation methods in applied mathematics.* Springer-Verlag, New York. [88]

[26] Milne-Thomson, L. M. (1948). *Theoretical aerodynamics.* MacMillan, London. [403]

[27] Milne-Thomson, L. M. (1949). *Theoretical hydrodynamics* (2nd edn). MacMillan, London. [155, 198, 402]

[28] Moffatt, H. K. and Tsinober, A. (1990). *Topological fluid mechanics. Proceedings of the IUTAM Symposium.* Cambridge University Press. [417]

[29] Ockendon, H. and Ockendon, J. R. (1995). *Viscous flow.* Cambridge University Press, New York. [161, 409]

[30] Ockendon, H. and Tayler, A. B. (1983). *Inviscid fluid flows.* Springer-Verlag, New York. [23, 53]

[31] Øksendal, B. (1992). *Stochastic differential equations* (3rd edn). Springer-Verlag, New York. [244, 395]

[32] Piaggio, H. T. H. (1946). *An elementary treatise on differential equations and their applications.* G. Bell & Sons Ltd, London. [360]

[33] Pólya, G. and Szegö, G. (1951). *Isoperimetric inequalities in mathematical physics.* Princeton University Press. [212]

[34] Polyanin, A. D. (2002). *Handbook of linear partial differential equations for engineers and scientists*, Chapman & Hall/CRC, Boca Raton. [174]

[35] Protter, M. H. and Weinberger, H. F. (1967). *Maximum principles in differential equations.* Prentice Hall, Englewood Cliffs, NJ. [249]

[36] Renardy, M. and Rogers, R. C. (1993). *An introduction to partial differential equations.* Springer-Verlag, New York. [68]

[37] Richards, I. and Youn, H. (1990). *Theory of distributions: a non-technical introduction.* Cambridge University Press. [97]

[38] Rubinstein, I. and Rubinstein, L. (1993). *Partial differential equations in classical mathematical physics.* Cambridge University Press. [329]

[39] Saffman, P. G. (1992). *Vortex dynamics*. Cambridge University Press, New York. [403]

[40] Sokolnikoff, I. S. (1956). *Mathematical theory of elasticity* (2nd edn). McGraw-Hill, New York. [128]

[41] Sparrow, E. M. and Cess, R. D. (1978). *Radiation heat transfer*. Hemisphere, New York. [155]

[42] Stakgold, I. (1998). *Green's functions and boundary value problems* (2nd edn). Wiley, New York. [97, 211]

[43] Tayler, A. B. (1986). *Mathematical models in applied mechanics*. Clarendon Press, Oxford. [vi]

[44] Titchmarsh, E. C. (1946). *Eigenfunction expansions associated with second-order differential equations*. Clarendon Press, Oxford. [107, 183]

[45] Van Dyke, M. (1982). *An album of fluid motion*. Parabolic Press, Stanford, CA. [377]

[46] Whittaker, E. T. (1937). *A treatise on the analytical dynamics of particles and rigid bodies with an introduction to the problem of three bodies* (4th edn). Cambridge University Press. [378]

[47] Wilmott, P., Howison, S. D. and Dewynne, J. N. (1995). *The mathematics of financial derivatives: a student introduction*. Cambridge University Press, New York. [244, 334, 335]

[48] Zauderer, E. (1989). *Partial differential equations of applied mathematics* (2nd edn). Wiley, Singapore. [184]

Index

Partial differential equations are used in mathematical models of a huge range of real-world phenomena, from electromagnetism to financial markets. This new edition of *Applied Partial Differential Equations* contains many new sections and exercises including: American options, transform methods, free surface flows, linear elasticity and complex characteristics.

New to this edition

- Numerous topical exercises
- Recent applications of PDEs
- Updates on transform methods (especially multi-dimensional Fourier transforms and the Radon transform), explicit representations of general solutions of the wave equation, bifurcations, the Wiener–Hopf method, free surface flows, American options, the Monge–Ampère equation, linear elasticity, and complex characteristics.

John Ockendon and **Sam Howison** are university lecturers working in the Oxford Centre for Industrial and Applied Mathematics, University of Oxford; **Andrew Lacey** is a professor at Heriot–Watt University; and **Alexander Movchan** is a professor at Liverpool University.

Contents

ISBN 0-19-852771-3
9 780198 527718

OXFORD
UNIVERSITY PRESS
www.oup.com