PDE & INTEGRAL EQUATIONS

(MTH3C14)

SELF LEARNING MATERIAL

III SEMESTER

M.Sc. MATHEMATICS
(2019 Admission)

UNIVERSITY OF CALICUT
SCHOOL OF DISTANCE EDUCATION
CALICUT UNIVERSITY P.O.
MALAPPURAM - 673 635, KERALA

190564

School of Distance Education
University of Calicut

Self Learning Material

III Semester
M.Sc. MATHEMATICS
(2019 Admission)

MTH3C14 : PDE & INTEGRAL EQUATIONS

Prepared by:

SREEHARI .T
Asst. Professor of Mathematics,
SDE, University of Calicut.

Dr. JICY N
Asst. Professor of Mathematics,
SDE, University of Calicut.

ASWINI N.K
Asst. Professor of Mathematics,
SDE, University of Calicut.

Scrutinized by:

SAVITHA K.N
Asst. Professor in Mathematics,
Mahatma Gandhi Government Arts College, Mahe.

DISCLAIMER

"The author(s) shall be solely responsible for the content and views expressed in this book".

Contents

Module 1 **1**

1 **1**

1.1 First-order equations 1

1.1.1 Introduction 1

1.1.2 Quasilinear equations 3

1.1.3 The method of characteristics 6

1.1.4 Examples of the characteristics method 13

1.1.5 The existence and uniqueness theorem 22

1.1.6 The Lagrange method 27

1.1.7 Conservation laws and shock waves 31

1.1.8 The eikonal equation 44

1.1.9 General nonlinear equations 48

1.2 Second-order linear equations in two independent variables 67

1.2.1 Introduction 67

1.2.2 Classification 68

1.2.3 Canonical form of hyperbolic equations 72

1.2.4 Canonical form of parabolic equations 76

1.2.5 Canonical form of elliptic equations 78

1.3 The one-dimensional wave equation 86

1.3.1 Introduction 86

1.3.2 Canonical form and general solution 86

1.3.3 The Cauchy problem and d'Alembert's formula . . 91

1.3.4 Domain of dependence and region of influence . . . 98

1.3.5 The Cauchy problem for the nonhomogeneous wave equation 105

Module 2 **123**

2 **124**

2.1 The method of separation of variables 124

2.1.1 Introduction 124

2.1.2 Heat equation: homogeneous boundary conditions 125

2.1.3 Separation of variables for the wave equation . . . 141

2.1.4 Separation of variables for nonhomogeneous equations 149

2.1.5 The energy method and uniqueness 152

2.1.6 Further applications of the heat equation 156

2.1.7 Wine cellars 164

2.2 Elliptic equations 169
2.2.1 Introduction 169
2.2.2 Basic properties of elliptic problems 170
2.2.3 The maximum principle 177
2.2.4 Applications of the maximum principle 182
2.2.5 Green's identities 183
2.2.6 The maximum principle for the heat equation . . . 186
2.2.7 Separation of variables for elliptic problems 190
2.2.8 Poisson's formula 213

Module 3 **219**

3 **220**

3.1 Introduction 220
3.2 Differential and integral equations 226
3.3 The Green's function 232
3.4 Fredholm equations with separable kernels 249
3.5 Illustrative example 254
3.6 Hilbert-Schmidt theory 258
3.7 Iterative methods 273
3.8 The Neumann series 284
3.9 Fredholm theory 289

Module 1

FIRST-ORDER EQUATIONS, SECOND ORDER LINEAR EQUATIONS AND THE ONE-DIMENSIONAL WAVE EQUATIONS

1.1 First-order equations

1.1.1 Introduction

A first-order PDE for an unknown function $u(x_1, x_2, \ldots, x_n)$ has the following general form:

$$F\left(x_1, x_2, \ldots, x_n, u, u_{x_1}, u_{x_2}, \ldots, u_{x_n}\right) = 0.$$

where F is a given function of $2n + 1$ variables. First-order equations appear in a variety of physical and engineering processes, such as the transport of material in a fluid flow and propagation of wavefronts in optics. Nevertheless they appear less frequently than second-order equations. For simplicity we shall limit the presentation in this chapter to functions in two variables. The reason for this is not just to simplify the algebra. As we shall soon observe, the solution method is based on the geometrical interpretation of u as a surface in an $(n+1)$-dimensional space.

We thus consider a surface in $\mathbb{R}^3$ whose graph is given by $u(x, y)$. The surface satisfies an equation of the form

$$F\left(x, y, u, u_x, u_y\right) = 0. \tag{1.1}$$

Equation (1.1) is still quite general. In many practical situations we deal with equations with a special structure that simplifies the solution process. Therefore we shall progress from very simple equations to more complex ones. There is a common thread to all types of equations - the geometrical approach. The basic idea is that since $u(x, y)$ is a surface in $\mathbb{R}^3$, and since the normal to the surface is given by the vector $(u_x, u_y, -1)$, the PDE (1.1) can be considered as an equation relating the surface to its normal (or alternatively its tangent plane). Indeed the main solution method will be a direct construction of the solution surface.

1.1.2 Quasilinear equations

We consider first a special class of nonlinear equations where the nonlinearity is confined to the unknown function u. The derivatives of u appear in the equation linearly. Such equations are called *quasilinear*. The general form of a quasilinear equation is

$$a(x, y, u)u_x + b(x, y, u)u_y = c(x, y, u). \tag{1.2}$$

An important special case of quasilinear equations is that of *linear equations*:

$$a(x, y)u_x + b(x, y)u_y = c_0(x, y)u + c_1(x, y), \tag{1.3}$$

where a, b, c_0, c_1 are given functions of (x, y).

Before developing the general theory for quasilinear equations, let us warm up with a simple example.

Example 1.1.1.

$$u_x = c_0 u + c_1 \tag{1.4}$$

In this example we set $a = 1, b = 0, c_0$ is a constant, and $c_1 = c_1(x, y)$. Since (1.4) contains no derivative with respect to the y variable, we can regard this variable as a parameter. Recall from the theory of ODEs that in order to obtain a unique solution we must supply an additional condition. There are many ways to supply additional conditions to a PDE. The natural condition for a first-order PDE is a curve lying on the solution surface. We shall refer to such a condition as an initial condition, and the problem will be called an initial value problem or a

Cauchy problem in honor of the French mathematician Augustin Louis Cauchy (1789-1857). For example, we can supplement (1.4) with the initial condition

$$u(0, y) = y. \tag{1.5}$$

Since we are actually dealing with an ODE, the solution is immediate:

$$u(x, y) = \mathrm{e}^{c_0 x} \left[\int_0^x \mathrm{e}^{-c_0 \xi} c_1(\xi, y) \mathrm{d}\xi + y \right]. \tag{1.6}$$

A basic approach for solving the general case is to seek special variables in which the equation is simplified (actually, similar to (1.4)). Before doing so, let us draw a few conclusions from this simple example.

(1) Notice that we integrated along the x direction (see Figure 1.1) from each point on the y axis where the initial condition was given, i.e. we actually solved an infinite set of ODEs.

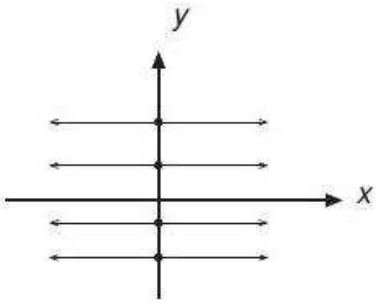

Figure 1.1: Integration of (1.4).

(2) Is there always a solution to (1.4) and an initial condition? At a first

sight the answer seems positive; we can write a general solution for (1.4) in the form

$$u(x,y) = \mathrm{e}^{c_0 x}\left[\int_0^x \mathrm{e}^{-c_0\xi} c_1(\xi,y)\mathrm{d}\xi + T(y)\right], \tag{1.7}$$

where the function $T(y)$ is determined by the initial condition. There are examples, however, where such a function does not exist at all! For instance, consider the special case of (1.4) in which $c_1 \equiv 0$. The solution (1.7) now becomes $u(x,y) = \mathrm{e}^{c_0 x}T(y)$. Replace the initial condition (1.5) with the condition

$$u(x,0) = 2x. \tag{1.8}$$

Now $T(y)$ must satisfy $T(0) = 2x\mathrm{e}^{-c_0 x}$, which is of course impossible.

(3) We have seen so far an example in which a problem had a unique solution, and an example where there was no solution at all. It turns out that an equation might have infinitely many solutions. To demonstrate this possibility, let us return to the last example, and replace the initial condition (1.5) by

$$u(x,0) = 2\mathrm{e}^{c_0 x}. \tag{1.9}$$

Now $T(y)$ should satisfy $T(0) = 2$. Thus every function $T(y)$ satisfying $T(0) = 2$ will provide a solution for the equation together with the initial condition. Therefore, (1.4) with $c_1 = 0$ has infinitely many solutions under the initial condition (1.9).

We conclude from Example 1.1.1 that the solution process must include the step of checking for existence and uniqueness.

1.1.3 The method of characteristics

We solve first-order PDEs by the method of *characteristics*. This method was developed in the middle of the nineteenth century by Hamilton. Hamilton investigated the propagation of light. Fermat proposed a unified principle, according to which light rays travel from a point A to a point B in an orbit that takes the least amount of time. Hamilton showed that this principle can serve as a foundation of a dynamical theory of rays. He thus derived an axiomatic theory that provided equations of motion for light rays. The main building block in the theory is a function that completely characterizes any given optical medium. Hamilton called it the *characteristic function*. He showed that Fermat's principle implies that his characteristic function must satisfy a certain first-order nonlinear PDE. Hamilton's characteristic function and characteristic equation are now called the *eikonal function* and eikonal equation after the Greek word $\epsilon\iota\kappa\omega\upsilon$ (or $\epsilon\iota\kappa O\upsilon$) which means "an image".

Hamilton discovered that the eikonal equation can be solved by integrating it along special curves that he called *characteristics*. Furthermore, he showed that in a uniform medium, these curves are exactly the straight light rays whose existence has been assumed since ancient times. Notice that although the eikonal equation is of first order, it is in fact fully nonlinear and not quasilinear.

We shall first develop the method of characteristics heuristically. Later we shall present a precise theorem that guarantees that, under suitable assumptions, the equation together with its associated condition has a unique solution. The characteristics method is based on 'knitting' the solution surface with a one-parameter family of curves that intersect a given curve in space. Consider the general linear equation (1.3), and

write the initial condition parameterically:

$$\Gamma = \Gamma(s) = (x_0(s), y_0(s), u_0(s)), \quad s \in I = (\alpha, \beta). \tag{1.10}$$

The curve Γ will be called the *initial curve*.

The linear equation (1.3) can be rewritten as

$$(a, b, c_0 u + c_1) \cdot (u_x, u_y, -1) = 0. \tag{1.11}$$

Since $(u_x, u_y, -1)$ is normal to the surface u, the vector $(a, b, c_0 u + c_1)$ is in the tangent plane. Hence, the system of equations

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t}(t) &= a(x(t), y(t)), \\ \frac{\mathrm{d}y}{\mathrm{d}t}(t) &= b(x(t), y(t)), \\ \frac{\mathrm{d}u}{\mathrm{d}t}(t) &= c(x(t), y(t)))u(t) + c_1(x(t), y(t)) \end{aligned} \tag{1.12}$$

defines spatial curves lying on the solution surface (conditioned so that the curves start on the surface). This is a system of first-order ODEs. They are called the *system of characteristic equations* or, for short, the *characteristic equations*. The solutions are called *characteristic curves* of the equation. Notice that equations (1.12) are autonomous, i.e. there is no explicit dependence upon the parameter t. In order to determine a characteristic curve we need an initial condition. We shall require the initial point to lie on the initial curve Γ. Since each curve $(x(t), y(t), u(t))$ emanates from a different point $\Gamma(s)$, we shall explicitly write the curves in the form $(x(t, s), y(t, s), u(t, s))$. The initial conditions are written as:

$$x(0,s) = x_0(s), \quad y(0,s) = y_0(s), \quad u(0,s) = u_0(s). \tag{1.13}$$

Notice that we selected the parameter t such that the characteristic curve is located on Γ when $t = 0$. One may, of course, select any other parameterization. We also notice that, in general, the parameterization $(x(t,s), y(t,s), u(t,s))$ represents a surface in $\mathbb{R}^3$.

One can readily verify that the method of characteristics applies to the quasilinear equation (1.2) as well. Namely, each point on the initial curve Γ is a starting point for a characteristic curve. The characteristic equations are now

$$\begin{aligned} x_t(t) &= a(x,y,u), \\ y_t(t) &= b(x,y,u), \\ u_t(t) &= c(x,y,u), \end{aligned} \tag{1.14}$$

supplemented by the initial condition

$$x(0,s) = x_0(s), \quad y(0,s) = y_0(s), \quad u(0,s) = u_0(s). \tag{1.15}$$

The problem consisting of (1.2) and initial conditions (1.15) is called the *Cauchy problem* for quasilinear equations.

The main difference between the characteristic equations (1.12) derived for the linear equation, and the set (1.14) is that in the former case the first two equations of (1.12) are independent of the third equation and of the initial conditions. We shall observe later the special role

played by the projection of the characteristic curves on the (x, y) plane. Therefore, we write (for the linear case) the equation for this

projection separately:

$$x_t = a(x, y), \quad y_t = b(x, y). \tag{1.16}$$

In the quasilinear case, this uncoupling of the characteristic equations is no longer possible, since the coefficients a and b depend upon u. We also point out that in the linear case, the equation for u is always linear, and thus it is guaranteed to have a global solution (provided that the solutions $x(t)$ and $y(t)$ exist globally).

To summarize the preliminary presentation of the method of characteristics, let us consult Figure 1.2. In the first step we identify the initial curve Γ. In the second step we select a point s on Γ and solve the characteristic equations (1.12) (or (1.14)), using the point we selected on Γ as an initial point. After performing these steps for all points on Γ we obtain a portion of the solution surface (also called the *integral surface*) that consists of the union of the characteristic curves. Philosophically speaking, one might say that the characteristic curves take with them an initial piece of information from Γ, and propagate it with

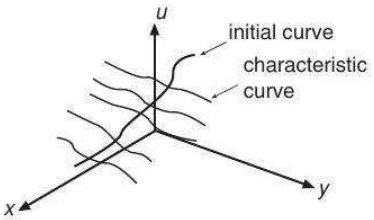

Figure 1.2: Sketch of the method of characteristics.

them. Furthermore, each characteristic curve propagates independently of the other characteristic curves.

Let us demonstrate the method for a very simple case.

Example 1.1.2. Solve the equation

$$u_x + u_y = 2$$

subject to the initial condition $u(x, 0) = x^2$.

The characteristic equations and the parametric initial conditions are

$$\begin{aligned} &x_t(t, s) = 1, \quad y_t(t, s) = 1, \quad u_t(t, s) = 2, \\ &x(0, s) = s, \quad y(0, s) = 0, \quad u(0, s) = s^2. \end{aligned}$$

It is a simple matter to solve for the characteristic curves:

$$x(t, s) = t + f_1(s), \quad y(t, s) = t + f_2(s), \quad u(t, s) = 2t + f_3(s).$$

Upon substituting into the initial conditions, we find

$$x(t, s) = t + s, \quad y(t, s) = t, \quad u(t, s) = 2t + s^2.$$

We have thus obtained a parametric representation of the integral surface. To find an explicit representation of the surface u as a function of x and y we need to invert the transformation $(x(t, s), y(t, s))$, and to express it in the form $(t(x, y), s(x, y))$, namely, we have to solve for (t, s) as functions of (x, y). In the current example the inversion is easy to

perform:

$$t = y, \quad s = x - y.$$

Thus the explicit representation of the integral surface is given by

$$u(x, y) = 2y + (x - y)^2.$$

This simple example might lead us to think that each initial value problem for a first-order PDE possesses a unique solution. But we have already seen that this is not the case. What, therefore, are the obstacles we might face? Is (1.2) equipped with initial conditions (1.13) well-posed? For simplicity we shall discuss in this chapter two aspects of well-posedness: existence and uniqueness. Thus the question is whether there exists a unique integral surface for (1.2) that contains the initial curve.

1. Notice that even if the PDE is linear, the characteristic equations are nonlinear! We know from the theory of ODEs that in general one can only establish local existence of a unique solution (assuming that the coefficients of the equation are smooth functions). In other words, the solutions of nonlinear ODEs might develop singularities within a short distance from the initial point even if the equation is very smooth. It follows that one can expect at most a local existence theorem for a first-order PDE, even if the PDE is linear.

2. The parametric representation of the integral surface might hide

further difficulties. We shall demonstrate this in the sequel by obtaining naive-looking parametric representations of singular surfaces. The difficulty lies in the inversion of the transformation from the plane (t, s) to the plane (x, y). Recall that the implicit function theorem implies that such a transformation is invertible if the Jacobian $J = \partial(x, y)/\partial(t, s) \neq 0$. But we observe that while the dependence of the characteristic curves on the variable t is derived from the PDE itself, the dependence on the variable s is derived from the initial condition. Since the equation and the initial condition do not depend upon each other, it follows that for any given equation there exist initial curves for which the Jacobian vanishes, and the implicit function theorem does not hold.

The functional problem we just described has an important geometrical interpretation. An explicit computation of the Jacobian at points located on the initial curve Γ, using the characteristic equations, gives

$$J = \frac{\partial x}{\partial t}\frac{\partial y}{\partial s} - \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} = \begin{vmatrix} a & b \\ (x_0)_s & (y_0)_s \end{vmatrix} = (y_0)_s\, a - (x_0)_s\, b, \quad (1.17)$$

where $(x_0)_s = \mathrm{d}x_0/\mathrm{d}s$. Thus the Jacobian vanishes at some point if and only if the vectors (a, b) and $((x_0)_s, (y_0)_s)$ are linearly dependent. Hence the geometrical meaning of a vanishing Jacobian is that the projection of Γ on the (x, y) plane is tangent at this point to the projection of the characteristic curve on that plane. As a rule, in order for a first-order quasilinear PDE to have a unique solution near the initial curve, we must have $J \neq 0$. This condition is called the *transversality condition*.

3. So far we have discussed local problems. One can also encounter global problems. For example, a characteristic curve might intersect the initial curve more than once. Since the characteristic equation is well-posed for a single initial condition, then in such a situation the solution will, in general, develop a singularity. We can think about this situation in the following way. Recall that a characteristic curve 'carries' with it along its orbit a charge of information from its intersection point with Γ. If a characteristic curve intersects Γ more than once, these two 'information charges' might be in conflict.

   A similar global problem is the intersection of the projection on the (x, y) plane of different characteristic curves with each other. Such an intersection is problematic for the same reason as the intersection of a characteristic curve with the initial curve. Each characteristic curve carries with it a different information charge, and a conflict might arise at such an intersection.

4. Another potential problem relates to a lack of uniqueness of the solution to the characteristic equation. We should not worry about this possibility if the coefficients of the equations are smooth (Lipschitz continuous, to be precise). But when considering a nonsmooth problem, we should pay attention to this issue. We shall demonstrate such a case below.

1.1.4 Examples of the characteristics method

Example 1.1.3. Solve the equation $u_x = 1$ subject to the initial condition $u(0, y) = g(y)$.

The characteristic equations and the associated initial conditions are

given by

$$x_t = 1, \quad y_t = 0, \quad u_t = 1, \tag{1.18}$$

$$x(0,s) = 0, \quad y(0,s) = s, \quad u(0,s) = g(s), \tag{1.19}$$

respectively. The parametric integral surface is $(x(t,s), y(t,s), u(t, s)) = (t, s, t+ g(s))$. It is easy to deduce from here the explicit solution $u(x,y) = x+g(y)$. On the other hand, if we keep the equation unchanged, but modify the initial conditions into $u(x,0) = h(x)$, the picture changes dramatically. In this case the parametric solution is

$$(x(t,s), y(t,s), u(t,s)) = (t+s, 0, t+h(s)).$$

Now, however, the transformation $(x(t,s), y(t,s))$ cannot be inverted. Geometrically speaking, the reason is simple: the projection of the initial curve is precisely the x axis, but this is also the projection of a characteristic curve. In the special case where $h(x) = x + c$ for some constant c, we obtain $u(t,s) = s + t + c$. Then it is not necessary to invert the mapping $(x(t,s), y(t,s))$, since we find at once $u = x+c+f(y)$ for every differentiable function $f(y)$ that vanishes at the origin. But for any other choice of h the problem has no solution at all.

We note that for the initial conditions $u(x,0) = h(x)$ we could have foreseen the problem through a direct computation of the Jacobian:

$$J = \begin{vmatrix} a & b \\ (x_0)_s & (y_0)_s \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix} = 0. \tag{1.20}$$

Whenever the Jacobian vanishes along an interval (like in the example

we are considering), the problem will, in general, have no solution at all. If a solution does exist, we shall see that this implies the existence of infinitely many solutions.

Because of the special role played by the projection of the characteristic curves on the (x, y) plane we shall use the term *characteristics* to denote them for short. There are several ways to compute the characteristics. One of them is to solve the full characteristic equations, and then to project the solution on the (x, y) plane. We note that the projection of a characteristic curve is given by the condition $s =$ constant. Substituting this condition into the equation $s = s(x, y)$ determines an explicit equation for the characteristics. An alternative method is valid whenever the PDE is linear. The linearity implies that the first two characteristic equations are independent of u. Thus they can be solved directly for the characteristics themselves without solving first for the parametric integral surface. Furthermore, since the characteristic equations are autonomous (i.e. they do not explicitly include the variable t), it follows that the equations for the characteristics can be written simply as the first-order ODE

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b(x, y)}{a(x, y)}$$

Example 1.1.4. The current example will be useful for solving linear equations of the form

$$a(x, y)u_x + b(x, y)u_y = 0$$

The equations for the characteristic curves

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = b, \quad \frac{\mathrm{d}u}{\mathrm{d}t} = 0$$

imply at once that the solution u is constant on the characteristics that are determined by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b(x,y)}{a(x,y)} \tag{1.21}$$

For instance, when $a = 1, b = \sqrt{-x}$ we obtain that u is constant along the lines $\frac{3}{2}y + (-x)^{3/2} =$ constant.

Example 1.1.5. Solve the equation $u_x + u_y + u = 1$, subject to the initial condition

$$u = \sin x, \text{ on } y = x + x^2, x > 0.$$

The characteristic equations and the associated initial conditions are given by

$$x_t = 1, \quad y_t = 1, \quad u_t + u = 1, \tag{1.22}$$

$$x(0,s) = s, \quad y(0,s) = s + s^2, \quad u(0,s) = \sin s, \tag{1.23}$$

respectively. Let us compute first the Jacobian along the initial curve:

$$J = \begin{vmatrix} 1 & 1 \\ 1 & 1+2s \end{vmatrix} = 2s. \tag{1.24}$$

Thus we anticipate a unique solution at each point where $s \neq 0$. Since we are limited to the regime $x > 0$ we indeed expect a unique

solution.

The parametric integral surface is given by

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

In order to invert the mapping $(x(t,s), y(t,s))$, we substitute the equation for x into the equation for y to obtain $s = (y-x)^{1/2}$. The sign of the square root was selected according to the condition $x > 0$. Now it is easy to find $t = x - (y-x)^{\frac{1}{2}}$, whence the explicit representation of the integral surface

$$u(x,y) = 1 - \left[1 - \sin(y-x)^{\frac{1}{2}}\right] \mathrm{e}^{-x+(y-x)^{\frac{1}{2}}}.$$

Notice that the solution exists only in the domain

$$D = \{(x,y) \mid 0 < x < y\} \cup \left\{(x,y) \mid x \leq 0 \text{ and } x + x^2 < y\right\},$$

and in particular it is not differentiable at the origin of the (x,y) plane. To see the geometrical reason for this, consult Figure 1.3. We see that the slope of characteristic passing through the origin equals 1 , which is exactly the slope of the projection of

the initial curve there. Namely, the transversality condition does not hold there (a fact we already expected from our computation of the Jacobian above). Indeed the violation of the transversality condition led to nonuniqueness of the solution near the curve

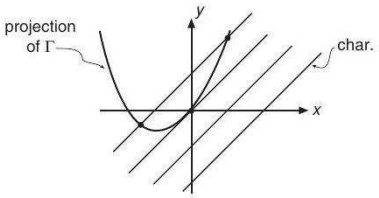

Figure 1.3: The characteristics and projection of Γ for Example 1.1.5

.

$$\left\{(x,y) \mid x < 0 \text{ and } y = x + x^2\right\}$$

which is manifested in the ambiguity of the sign of the square root.

Example 1.1.6. Solve the equation $-yu_x + xu_y = u$ subject to the initial condition $u(x,0) = \psi(x)$.

The characteristic equations and the associated initial conditions are given by

$$x_t = -y, \quad y_t = x, \quad u_t = u, \tag{1.25}$$

$$x(0,s) = s, \quad y(0,s) = 0, \quad u(0,s) = \psi(s). \tag{1.26}$$

Let us examine the transversality condition:

$$J = \begin{vmatrix} 0 & s \\ 1 & 0 \end{vmatrix} = -s. \tag{1.27}$$

Thus we expect a unique solution (at least locally) near each point on the initial curve, except, perhaps, the point $x = 0$.

The solution of the characteristic equations is given by

$$\begin{aligned}&(x(t,s),y(t,s),u(t,s))\\&\quad=\left(f_1(s)\cos t+f_2(s)\sin t, f_1(s)\sin t-f_2(s)\cos t, \mathrm{e}^t f_3(s)\right).\end{aligned}$$

Substituting the initial condition into the solution above leads to the parametric integral surface

$$(x(t,s),y(t,s),u(t,s))=\left(s\cos t, s\sin t, \mathrm{e}^t\psi(s)\right).$$

Isolating s and t we obtain the explicit representation

$$u(x,y)=\psi\left(\sqrt{x^2+y^2}\right)\exp\left[\arctan\left(\frac{y}{x}\right)\right].$$

It can be readily verified that the characteristics form a one-parameter family of circles around the origin (see Figure 1.4). Therefore, each one of them intersects the projection of the initial curve (the x axis) twice. We also saw that the Jacobian vanishes at the origin. So how is it that we seem to have obtained a unique solution? The mystery is easily re-

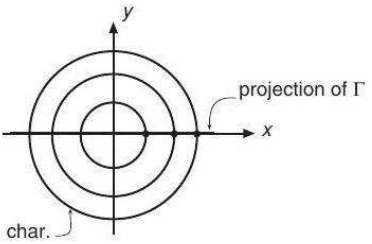

Figure 1.4: The characteristics and projection of Γ for Example 1.1.6.

solved by observing that in choosing the positive sign for the square root in the argument of ψ, we effectively reduced the solution to the ray $\{x > 0\}$. Indeed, in this region a characteristic intersects the projection of the initial curve only once.

Example 1.1.7. Solve the equation $u_x + 3y^{2/3}u_y = 2$ subject to the initial condition $u(x, 1) = 1 + x$.

The characteristic equations and the associated initial conditions are given by

$$x_t = 1, \quad y_t = 3y^{2/3}, \quad u_t = 2, \tag{1.28}$$

$$x(0, s) = s, \quad y(0, s) = 1, \quad u(0, s) = 1 + s. \tag{1.29}$$

In this example we expect a unique solution in a neighborhood of the initial curve since the transversality condition holds:

$$J = \begin{vmatrix} 1 & 3 \\ 1 & 0 \end{vmatrix} = -3 \neq 0. \tag{1.30}$$

The parametric integral surface is given by

$$x(t, s) = s + t, \quad y(t, s) = (t + 1)^3, \quad u(t, s) = 2t + 1 + s.$$

Before proceeding to compute an explicit solution, let us find the characteristics. For this purpose recall that each characteristic curve passes through a specific s value. Therefore, we isolate t from the equation for x, and substitute it into the expression for y. We obtain $y = (x + 1 - s)^3$, and, thus, for each fixed s this is an equation for a

characteristic.

A number of characteristics and their intersection with the projection of the initial curve $y = 1$ are sketched in Figure 1.5. While the picture indicates no problems, we were not careful enough in solving the characteristic equations, since the function $y^{2/3}$ is not Lipschitz continuous at the origin. Thus the characteristic equations might not have a unique solution there! In fact, it can be easily verified that $y = 0$ is also a solution of $y_t = 3y^{2/3}$. But, as can be seen from Figure 1.5, the well behaved characteristics near the projection of the initial curve $y = 1$ intersect at some point the extra characteristic $y = 0$. Thus we can anticipate irregular behavior near $y = 0$. Inverting the mapping $(x(t, s), y(t, s))$ we obtain

$$t = y^{1/3} - 1, s = x + 1 - y^{1/3}.$$

Hence the explicit solution to the PDE is $u(x, y) = x + y^{1/3}$, which is indeed singular on the x axis.

Example 1.1.8. Solve the equation $(y + u)u_x + yu_y = x - y$ subject to the initial conditions $u(x, 1) = 1 + x$.

This is an example of a quasilinear equation. The characteristic equa-

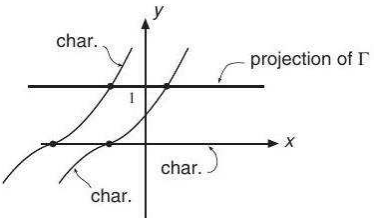

Figure 1.5: Self-intersection of characteristics.

tions and the initial data are:

$$\begin{array}{lll} (i) \quad x_t = y + u, & (ii) \quad y_t = y, & (iii) \quad u_t = x - y, \\ \quad x(0,s) = s, & \quad y(0,s) = 1, & \quad u(0,s) = 1 + s \ . \end{array}$$

Let us examine the transversality condition. Notice that while u is yet to be found, the transversality condition only involves the values of u on the initial curve Γ. It is easy to verify that on Γ we have $a = 2 + s, \quad b = 1$. It follows that the tangent to the characteristic has a nonzero component in the direction of the y axis. Thus it is nowhere tangent to the projection of the initial curve (the x axis, in this case). Alternatively, we can compute the Jacobian directly:

$$J = \begin{vmatrix} 2+s & 1 \\ 1 & 0 \end{vmatrix} = -1 \neq 0. \tag{1.31}$$

We conclude that there exists an integral surface at least at the vicinity of Γ. From the characteristic equation (ii) and the associated initial condition we find $y(t,s) = \mathrm{e}^t$. Adding the characteristic equations (i) and (iii) we get $(x+u)_t = x + u$. Therefore, $u + x = (1 + 2s)\mathrm{e}^t$. Returning to (i) we obtain $x(t,s) = (1+s)\mathrm{e}^t - \mathrm{e}^{-t}$ and $u(t,s) = s\mathrm{e}^t + \mathrm{e}^{-t}$. Observing that $x - y = s\mathrm{e}^t - \mathrm{e}^{-t}$, we finally get $u = 2/y + (x - y)$. The solution is not global (it becomes singular on the x axis), but it is well defined near the initial curve.

1.1.5 The existence and uniqueness theorem

We shall summarize the discussion on linear and quasilinear equations into a general theorem. For this purpose we need the following definition.

Definition 1. Consider a quasilinear equation (1.2) with initial condi-

tions (1.15) defining an initial curve for the integral surface. We say that the equation and the initial curve satisfy the transversality condition at a point s on Γ, if the characteristic emanating from the projection of $\Gamma(s)$ intersects the projection of Γ nontangentially, i.e.

$$J|_{t=0} = x_t(0,s)y_s(0,s) - y_t(0,s)x_s(0,s) = \begin{vmatrix} a & b \\ (x_0)_s & (y_0)_s \end{vmatrix} \neq 0.$$

Theorem 1.1.9. *Assume that the coefficients of the quasilinear equation (*1.2*) are smooth functions of their variables in a neighborhood of the initial curve (*1.15*). Assume further that the transversality condition holds at each point s in the interval $(s_0 - 2\delta, s_0 + 2\delta)$ on the initial curve. Then the Cauchy problem (*1.2*), (*1.15*) has a unique solution in the neighborhood $(t,s) \in (-\epsilon,\epsilon) \times (s_0 - \delta, s_0 + \delta)$ of the initial curve. If the transversality condition does not hold for an interval of s values, then the Cauchy problem (*1.2*), (*1.15*) has either no solution at all, or it has infinitely many solutions.*

Proof. The existence and uniqueness theorem for ODEs, applied to (1.14) together with the initial data (1.14), guarantees the existence of a unique characteristic curve for each point on the initial curve. The family of characteristic curves forms a parametric representation of a surface. The transversality condition implies that the parametric representation provides a smooth surface. Let us verify now that the surface thus constructed indeed satisfies the PDE (1.2). We write

$$\tilde{u} = \tilde{u}(x,y) = u(t(x,y), s(x,y))$$

and compute

$$\begin{aligned} a\tilde{u}_x + b\tilde{u}_y &= a\left(u_t t_x + u_s s_x\right) + b\left(u_t t_y + u_s s_y\right) \\ &= u_t\left(at_x + bt_y\right) + u_s\left(as_x + bs_y\right). \end{aligned}$$

But the characteristic equations and the chain rule imply

$$1 = t_t = at_x + bt_y, \quad 0 = s_t = as_x + bs_y.$$

Hence $a\tilde{u}_x + b\tilde{u}_y = u_t = c$, i.e. $\tilde{u}$ satisfies (1.2).

To show that there are no further integral surfaces, we prove that the characteristic curves we constructed must lie on an integral surface. Since the characteristic curve starts on the integral surface, we only have to show that it remains there. This is intuitively clear, since the characteristic curve is, by definition, orthogonal at every point to the surface normal. On the other hand, clearly for a curve starting on some surface to leave the surface, its tangent must at some point have a nonzero projection on the normal to the surface. This simple geometrical reasoning can be supported through an explicit computation; for this purpose we write a given integral surface in the form $u = f(x, y)$. Let $(x(t), y(t), u(t))$ be a characteristic curve. We assume $u(0) = f(x(0), y(0))$. Define the function

$$\Psi(t) = u(t) - f(x(t), y(t)).$$

Differentiating by t we write

$$\Psi_t = u_t - f_x(x,y)x_t - f_y(x,y)y_t.$$

Substituting the system (1.14) into the above equations for Ψ_t we obtain

$$\Psi_t = c(x,y,\Psi+f) - f_x(x,y)a(x,y,\Psi+f) - f_y(x,y)b(x,y,\Psi+f) \quad (1.32)$$

But the initial condition implies $\Psi(0) = 0$. It is easy to check (using (1.2) that $\Psi(t) \equiv 0$ solves the ODE (1.32). Since that equation has smooth coefficients, it has a unique solution. Thus $\Psi \equiv 0$ is the only solution, and the curve $(x(t), y(t), u(t))$ indeed lies on the integral surface. Therefore the integral surface we constructed earlier through the parametric representation induced by the characteristic equations is unique.

When the transversality condition does not hold along an interval of s values, the characteristic there is the same as the projection of Γ. If the solution of the characteristic equation is a curve that is not identical to the initial curve, then the tangent (vector) to the initial curve at some point cannot be at that point tangential to any integral surface. In other words, the initial condition contradicts the equation and thus there can be no solution to the Cauchy problem. If, on the other hand, the characteristic curve agrees with the initial curve at that point, there are infinitely many ways to extend it into a compatible integral surface that contains it. Therefore, in this case we have infinitely many solutions to the Cauchy problem. We now present a method for constructing this family of solutions. Select an arbitrary point $P_0 = (x_0, y_0, u_0)$ on Γ.

Construct a new initial curve Γ', passing through P_0, which is not tangent to Γ at P_0. Solve the new Cauchy problem consisting of (1.2) with Γ' as initial curve. Since, by construction, the transversality condition holds now, the first part of the theorem guarantees a unique solution. Since there are infinitely many ways of selecting such an initial curve Γ', we obtain infinitely many solutions. □

The following simple example demonstrates the case where the transversality condition fails along some interval.

Example 1.1.10. Consider the Cauchy problem

$$u_x + u_y = 1, \quad u(x, x) = x.$$

Show that it has infinitely many solutions.

The transversality condition is violated identically. However the characteristic direction is $(1, 1, 1)$, and so is the direction of the initial curve. Hence, the initial curve is itself a characteristic curve. Thus there exist infinitely many solutions. To find these solutions, set the problem

$$u_x + u_y = 1, \quad u(x, 0) = f(x),$$

for an arbitrary f satisfying $f(0) = 0$. The solution is easily found to be $u(x, y) = y + f(x - y)$.

Notice that the Cauchy problem

$$u_x + u_y = 1, \quad u(x, x) = 1,$$

on the other hand, is not solvable. To see this observe that the

transversality condition fails again, but now the initial curve is not a characteristic curve. Thus there is no solution.

Remark 1.1.1. There is one additional possibility not covered by Theorem 1.1.9. This is the case where the transversality condition does not hold on isolated points (as was indeed the case in some of the preceding examples). It is difficult to formulate universal statements here. Instead, each such case has to be analyzed separately.

1.1.6 The Lagrange method

First-order quasilinear equations were in fact studied by Lagrange even before Hamilton. Lagrange developed a solution method that is also geometric in nature, albeit less general than Hamilton's method. The main advantage of Lagrange's method is that it provides general solutions for the equation, regardless of the initial data.

Let us reconsider (1.14). The set of all solutions to this system forms a two parameter set of curves. To justify this assertion, notice that since the system (1.14) is autonomous, it is equivalent to the system

$$y_x = b(x, y, u)/a(x, y, u), \quad u_x = c(x, y, u)/a(x, y, u) \tag{1.33}$$

Since (1.33) is a system of two first-order ODEs in the (y, u) plane, where x is a parameter, it follows that the set of solutions is determined by two initial conditions.

Lagrange assumed that the two-parameter set of solution curves for (1.14) can be represented by the intersection of two families of integral

surfaces

$$\psi(x, y, u) = \alpha, \quad \phi(x, y, u) = \beta. \tag{1.34}$$

When we vary the parameters α and β we obtain (through intersecting the surfaces ψ and ϕ) the two-parameter set of curves that are generated by the intersection. Recall that a solution surface of (1.2) passing through an initial curve is obtained from a one-parameter family of curves solving the characteristic equation (1.14). Each such one-parameter subfamily describes a curve in the parameter space (α, β). Since such a curve can be expressed in the form $F(\alpha, \beta) = 0$, it follows that every solution of (1.2) and (1.15) is given by

$$F(\psi(x, y, u), \phi(x, y, u)) = 0. \tag{1.35}$$

Since the surfaces ψ and ϕ were determined by the equation itself, (1.35) defines a general solution to the PDE. When we solve for a particular initial curve, we just have to use this curve to determine the specific functional form of F associated with that initial curve.

We are still left with one "little" problem: how to find the surfaces ψ and ϕ. In the theory of ODEs one solves first-order equations by the method of integration factors. While this method is always feasible in theory, it involves great technical difficulties. In fact, it is possible to find integration factors only in special cases. In a sense, the Lagrange method is a generalization of the integration factor method for ODEs, as we have to find solution surfaces for the two first-order ODEs (1.33). Hence it is not surprising that the method is applicable only in special cases. We proceed by introducing a method for computing the surfaces

ψ and ϕ, and then apply the method to a specific example.

Example 1.1.11. Recall that by definition, the surfaces $\psi = \alpha, \phi = \beta$ contain the characteristic curves. Assume there exist two independent vector fields $\vec{P}_1 = (a_1, b_1, c_1)$ and $\vec{P}_2 = (a_2, b_2, c_2)$ (i.e. they are nowhere tangent to each other) that are both orthogonal to the vector $\vec{P} = (a, b, c)$ (the vector defining the characteristic equations). This means $aa_1 + bb_1 + cc_1 = 0 = aa_2 + bb_2 + cc_2$. Let us assume further that the vector fields $\vec{P}_1$ and $\vec{P}_2$ are exact, i.e. $\nabla \times \vec{P}_1 = 0 = \nabla \times \vec{P}_2$. This implies that there exist two potentials ψ and ϕ satisfying $\nabla\psi = \vec{P}_1$ and $\nabla\phi = \vec{P}_2$. By construction it follows that $\mathrm{d}\psi = \nabla\psi \cdot \vec{P} = 0 = \nabla\phi \cdot \vec{P} = \mathrm{d}\phi$, namely, ψ and ϕ are constant on every characteristic curve, and form the requested two-parameter integral surfaces.

Let us apply this method to find the general solution to the equation

$$-yu_x + xu_y = 0. \tag{1.36}$$

The characteristic equations are

$$x_t = -y, \quad y_t = x, \quad u_t = 0.$$

In this example $\vec{P} = (-y, x, 0)$. It is easy to guess orthogonal vector fields $\vec{P}_1 = (x, y, 0)$ and $\vec{P}_2 = (0, 0, 1)$. The reader can verify that they are indeed exact vector fields. The associated potentials are

$$\psi(x, y, u) = \frac{1}{2}\left(x^2 + y^2\right), \quad \phi(x, y, u) = (0, 0, u).$$

Therefore, the general solution of (1.36) is given by

$$F\left(x^2+y^2, u\right) = 0 \tag{1.37}$$

or

$$u = g\left(x^2+y^2\right) \tag{1.38}$$

To find the specific solution of (1.36) that satisfies a given initial condition, we shall use that condition to eliminate g. For example, let us compute the solution of (1.36) satisfying $u(x,0) = \sin x$ for $x > 0$. Substituting the initial condition into (1.38) yields $g(\xi) = \sin\sqrt{\xi}$; hence $u(x,y) = \sin\sqrt{x^2+y^2}$ is the required solution.

While the Lagrange method has an advantage over the characteristics method, since it provides a *general* solution to the equation, valid for all initial conditions, it also has a number of disadvantages.

(1) We have already explained that ψ and ϕ can only be found under special circumstances. Many tricks for this purpose have been developed since Lagrange's days, yet, only a limited number of equations can be solved in this way.

(2) It is difficult to deduce from the Lagrange method any potential problems arising from the interaction between the equation and the initial data.

(3) The Lagrange method is limited to quasilinear equations. Its generalization to arbitrary nonlinearities is very difficult. On the other hand, as we shall soon see, the characteristics method can be naturally extended to a method that is applicable to general nonlinear PDEs.

1.1.7 Conservation laws and shock waves

The existence theorem for quasilinear equations only guarantees (under suitable conditions) the existence of a local solution. Nevertheless, there are cases of interest where we need to compute the solution of a physical problem beyond the point where the solution breaks down. In this section we shall discuss such a situation. For simplicity we shall perform the analysis in some detail for a canonical prototype of quasilinear equations given by

$$u_y + uu_x = 0. \tag{1.39}$$

This equation plays an important role in hydrodynamics. It models the flow of mass with concentration u, where the speed of the flow depends on the concentration. The variable y has the physical interpretation of time. We shall show that the solutions to this equation often develop a special singularity that is called a *shock wave*. In hydrodynamics the equation is called the *Euler equation*. Towards the end of the section we shall generalize the analysis that is performed for (1.39) to a larger family of equations, and in particular, we shall apply the theory to study traffic flow.

As a warm-up we start with the simple linear equation

$$u_y + cu_x = 0. \tag{1.40}$$

The difference between this equation and (1.39), is that in (1.40) the flow speed is given by the positive constant c. The initial condition

$$u(x,0) = h(x) \tag{1.41}$$

will be used for both equations. Solving the characteristic equations for the linear equation (1.40) we get

$$(x, y, u) = (s + ct, t, h(s)).$$

Eliminating s and t yields the explicit solution $u = h(x - cy)$. The solution implies that the initial profile does not change; it merely moves with speed c along the positive x axis, namely, we have a fixed wave, moving with a speed c while preserving the initial shape.

Euler's equation (1.39) is solved similarly. The characteristic equations are

$$x_t = u, \quad y_t = 1, \quad u_t = 0,$$

and their solution is

$$(x, y, u) = (s + h(s)t, t, h(s)),$$

where we used the parameterization $x(0, s) = s, y(0, s) = 0, u(0, s) = h(s)$ for the initial data. Therefore, the solution of the PDE is

$$u = h(x - uy) \tag{1.42}$$

except that this time this solution is actually implicit. In order to analyze this solution further we eliminate the y variable from the equa-

tions for the characteristics (the projection of the characteristic curves on the (x, y) plane):

$$x = s + h(s)y. \tag{1.43}$$

The third characteristic equation implies that for each fixed s, i.e. along each characteristic, u preserves its initial value $u = h(s)$. The other characteristic equations imply, then, that the characteristics are straight lines.

Since different characteristics have different slopes that are determined by the initial values of u, they might intersect. Such an intersection has an obvious physical interpretation that can be seen from (1.43): The initial data $h(s)$ determine the speed of the characteristic emanating from a given s. Therefore, if a characteristic leaving the point s_1 has a higher speed than a characteristic leaving the point s_2, and if $s_1 < s_2$, then after some (positive) time the faster characteristic will overtake the slower one.

As we explained above, the solution is not well defined at points where characteristic curves intersect. To see the resulting difficulty from an algebraic perspective, we differentiate the implicit solution with respect to x to get $u_x = h'(1 - yu_x)$, implying

$$u_x = \frac{h'}{1 + yh'} \tag{1.44}$$

Recalling that physically the variable y stands for time, we consider the ray $y > 0$. We conclude that the solution's derivative blows up at the critical time

$$y_c = -\frac{1}{h'(s)}. \tag{1.45}$$

Hence the classical solution is not defined for $y > y_c$. This conclusion is consistent with the heuristic physical interpretation presented above. Indeed a necessary condition for a singularity formation is that $h'(s) < 0$ at least at one point, such that a faster characteristic will start from a point behind a slower characteristic. If $h(s)$ is never decreasing, there will be no singularity; however, such data are exceptional. Observe that the solution becomes singular at the first time y that satisfies (1.45); such a value is achieved for the value s, where $h'(s)$ is minimal.

Equation (1.39) arises in the investigation of a fundamental physical problem. Thus we cannot end our analysis when a singularity forms. In other words, while the solution becomes singular at the critical time (1.45), the fluid described by the equations keeps flowing unaware of our mathematical troubles! Therefore we must find a means of extending the solution beyond y_c.

Extending singular solutions is not a simple matter. There are several ways to construct such extensions, and we must select a method that conforms with fundamental physical principles. The basic idea is to define a new problem. This new problem is formulated so as to be satisfied by each classical solution of the Euler equation, and such that each continuously differentiable solution of the new problem will satisfy the Euler equation. Yet, the new problem will also have nonsmooth solutions. A solution of the new extended problem is called a weak solution, and the new problem itself is called the weak formulation of the original PDE. We shall see that sometimes there exist more than one weak solution, and this will require upgrading the weak formulation

to include a selection principle.

We choose to formulate the weak problem by replacing the differential equation with an integral balance. Integral balance is more fundamental, and can only be transformed into a differential relation when the functions involved are sufficiently smooth. To apply the integral balance method we rewrite (1.39) in the form

$$\partial_y u + \frac{1}{2}\frac{\partial}{\partial x}\left(u^2\right) = 0, \tag{1.46}$$

and integrate (with respect to x, and for a fixed y) over an arbitrary interval $[a, b]$ to obtain

$$\partial_y \int_a^b u(\xi, y)\mathrm{d}\xi + \frac{1}{2}\left[u^2(b, y) - u^2(a, y)\right] = 0. \tag{1.47}$$

It is clear that every solution of the PDE satisfies the integral relation (1.47) as well. Also, since a and b are arbitrary, any function $u \in C^1$ that satisfies (1.47) would also satisfy the PDE. Nevertheless, the integral balance is also well defined for functions not in C^1; actually, (1.47) is even defined for functions with finitely many discontinuities.

We now demonstrate the construction of a weak solution that is a smooth function (continuously differentiable) except for discontinuities along a curve $x = \gamma(y)$. Since the solution is smooth on both sides of γ, it satisfies the equation there. It remains to compute γ. For this purpose we write the weak formulation in the form

$$\partial_y \left[\int_a^{\gamma(y)} u(\xi, y)\mathrm{d}\xi + \int_{\gamma(y)}^b u(\xi, y)\mathrm{d}\xi\right] + \frac{1}{2}\left[\left(u^2(b, y) - u^2(a, y)\right]\right. = 0.$$

Differentiating the integrals with respect to y and using the PDE itself leads to

$$\gamma_y(y)u^- - \gamma_y(y)u^+ - \frac{1}{2}\left[\int_a^{\gamma(y)} \left(u^2(\xi,y)\right)_\xi \,\mathrm{d}\xi + \int_{\gamma(y)}^b \left(u^2(\xi,y)\right)_\xi \,\mathrm{d}\xi\right] + \frac{1}{2}\left[u^2(b,y) - u^2(a,y)\right] = 0$$

Here we used $u^-$and $u^+$to denote the values of u when we approach the curve γ from the left and from the right, respectively. Performing the integration we obtain

$$\gamma_y(y) = \frac{1}{2}\left(u^- + u^+\right), \tag{1.48}$$

namely, the curve γ moves at a speed that is the average of the speeds on the left and right ends of it.

Example 1.1.12. Consider the Euler equation (1.39) with the initial conditions

$$u(x,0) = h(x) = \begin{cases} 1 & x \le 0 \\ 1 - x/\alpha & 0 < x < \alpha \\ 0 & x \ge \alpha \end{cases} \tag{1.49}$$

Since $h(x)$ is not monotone increasing, the solution will develop a singularity at some finite (positive) time. Formula (1.45) implies $y_c = \alpha$. For all $y < \alpha$ the (smooth) solution is given by

$$u(x,y) = \begin{cases} 1 & x \le y, \\ \frac{x-\alpha}{y-\alpha} & y < x < \alpha, \\ 0 & x \ge \alpha. \end{cases} \tag{1.50}$$

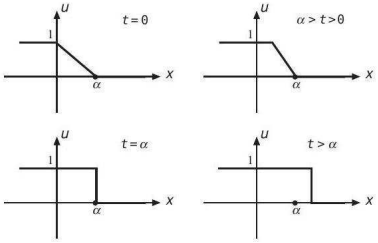

Figure 1.6: Several snapshots in the development of a shock wave.

After the critical time y_c when the solution becomes singular we need to define a weak solution. We seek a solution with a single discontinuity. Formula (1.48) implies that the discontinuity moves with a speed $\frac{1}{2}$. Therefore the following weak solution is compatible with the integral balance even for $y > \alpha$:

$$u(x,y) = \begin{cases} 1 & x < \alpha + \frac{1}{2}(y-\alpha), \\ 0 & x > \alpha + \frac{1}{2}(y-\alpha). \end{cases} \tag{1.51}$$

The solution thus constructed has the structure of a moving jump discontinuity. It describes a step function moving at a constant speed. Such a solution is called a *shock wave*. Several snapshots of the formation and propagation of a shock wave are depicted in Figure 1.6.

Strictly speaking, the solution is not continuously differentiable even at time $y = 0$; however, this is a minor complication, since it can be

shown that the formula for the classical solution is valid even when the derivative of the initial data has finitely many discontinuities as long as it is bounded.

Example 1.1.13. We now consider the opposite case where the initial data are increasing:

$$u(x,0) = \begin{cases} 0 & x \le 0, \\ x/\alpha & 0 < x < \alpha, \\ 1 & x \ge \alpha. \end{cases} \tag{1.52}$$

Since this time $h' \ge 0$, there is no critical (positive) time where the characteristics intersect. On the contrary, the characteristics diverge. This situation is called an *expansion wave*, in contrast to the wave in the previous example which is called a *compression wave*. We use the classical solution formula to obtain

$$u(x,y) = \begin{cases} 0 & x \le 0, \\ \frac{x}{\alpha+y} & 0 < x < \alpha + y, \\ 1 & x \ge \alpha + y. \end{cases} \tag{1.53}$$

It is useful to consider for both examples the limiting case where $\alpha \to 0$. In Example 1.1.12 the initial data are the same as the shock weak solution (1.51), and therefore this solution is already valid at $y = 0$. In contrast, in Example 1.1.13 the characteristics expand, the singularity is smoothed out at once, and the solution is

$$u(x,y) = \begin{cases} 0 & x \le 0, \\ \frac{x}{y} & 0 < x < y, \\ 1 & x \ge y. \end{cases} \tag{1.54}$$

We notice, though, that we could in principle write in the expansion wave case a weak solution that has a shock wave structure:

$$u(x,y) = \begin{cases} 0 & x < \alpha + \frac{1}{2}(y-\alpha), \\ 1 & x > \alpha + \frac{1}{2}(y-\alpha). \end{cases} \tag{1.55}$$

We see that the weak formulation by itself does not have a unique solution! Additional arguments are needed to pick out the correct solution among the several options. In the case we consider here it is intuitively clear that the shock solution (1.55) is not adequate, since slower characteristics starting from the ray $x < 0$ cannot overtake the faster characteristics that start from the ray $x > 0$. Another, more physical, approach to this problem will be described now.

The theory of weak solutions and shock waves is quite difficult. We therefore present essentially the same ideas we developed above from a somewhat different perspective. Instead of looking at the specific canonical equation (1.46) with general initial conditions, we look at a more general PDE with canonical initial conditions. Specifically we consider the following first-order quasilinear PDE

$$u_y + \frac{\partial}{\partial x} F(u) = 0. \tag{1.56}$$

Equations of this kind are called *conservation laws*. Let F denotes

flux. The flux is $F = \frac{1}{2}u^2$, and u is typically interpreted as mass density.

Equation (1.56) is supplemented with the initial condition

$$u(x,0) = \begin{cases} u^- & x < 0, \\ u^+ & x > 0. \end{cases} \tag{1.57}$$

To write the weak formulation for (1.56), we assume that the solution takes the shape of a shock wave of the form

$$u(x,y) = \begin{cases} u^- & x < \gamma(y), \\ u^+ & x > \gamma(y). \end{cases} \tag{1.58}$$

It remains therefore to find the shock orbit $x = \gamma(y)$. We find γ by integrating (1.56) with respect to x along the interval (x_1, x_2), with $x_1 < \gamma, x_2 > \gamma$. Taking (1.58) into account, we get

$$\frac{\partial}{\partial y}\left\{[x_2 - \gamma(y)]\, u^+ + [\gamma(y) - x_1]\, u^-\right\} = F\left(u^+\right) - F\left(u^-\right) \tag{1.59}$$

The last equation implies

$$\gamma_y(y) = \frac{F\left(u^+\right) - F\left(u^-\right)}{u^+ - u^-} := \frac{[F]}{[u]}, \tag{1.60}$$

where we used the notation $[.]$ to denote the change (jump) of a quantity across the shock.

Conservation laws appear in many areas of continuum mechanics (including hydrodynamics, gas dynamics, combustion, etc.), where the

jump equation (1.60) is called the Rankine-Hugoniot condition.

Notice that in the special case of $F = \frac{1}{2}u^2$ that we considered earlier, the rule (1.60) reduces to (1.48). We also point out that we integrated (1.56) along an arbitrary finite interval, although the values of x_1 and x_2 do not appear in the final conclusion. The reason for introducing this artificial interval is that the integral of (1.56) over the real line is not bounded. Since we interpret u as a density of a physical quantity (such as mass), our model is really artificial, and a realistic model will have to take into account the effects of finite boundaries.

Our analysis of the general conservation law (1.56) is not yet complete. From our study of Example 1.1.13 we expect that shock would only occur if the characteristics collide. In the case of general conservation laws, this condition is expressed as:

The entropy condition Characteristics must enter the shock curve, and are not allowed to emanate from it.

The motivation for the entropy condition is rooted in gas dynamics and the second law of thermodynamics. In order not to stray too much away from the theory of PDEs, we shall give a heuristic reasoning, based on the interpretation of entropy as minus the amount of "information" stored in a given physical system. We thus phrase the second law of thermodynamics as stating that in a closed system information is only lost as time y increases, and cannot be created. Now, we have shown that characteristics carry with them information on the solution of a first-order PDE. Therefore the emergence of a characteristic from a shock is interpreted as a creation of information which should be forbidden. To give the entropy condition an algebraic form, we write the conservation law (1.56) as $u_y + F_u u_x = 0$. Therefore the characteristic speed is given by F_u, and the entropy condition can be expressed as

$$F_u\left(u^-\right) > \gamma_y > F_u\left(u^+\right). \tag{1.61}$$

Applying this rule to the special case $F(u) = \frac{1}{2}u^2$ and using (1.60) we obtain that the shock solution is valid only if $u^- > u^+$, a conclusion we reached earlier from different considerations.

The theory of conservation laws has a nice application to the real-world problem of traffic flow. We therefore end this section by a qualitative analysis of this problem. Consider the flow of cars along one direction in a road. Although cars are discrete entities, we model them as a continuum, and denote by $u(x, y)$ the car density at a point x and time y.

A great deal of research has been devoted to the question of how to model the flux term $F(u)$. Clearly the flux is very low in bumper-to-bumper traffic, where each car barely moves. It may be surprising at first sight, but the flux is also low when the traffic is very light. In this case drivers tend to drive fast, and maintain a large distance between each other (at least this is what they ought to do when they drive fast...). Therefore the flux, which counts the total number of cars crossing a given point per time, is low. If we assume that a car occupies on average a length of 5 m, then the highest density is $u_{\mathrm{b}} = 200$ cars/ km. It was found experimentally that the maximal flux is about $F_{\max} =$ 1500 cars/hour, and is achieved at a speed of about 35 km/ hour (20 miles/hour). Therefore the optimal density (if one wants to maximize the flux) is $u_{\max} \sim 43$ cars/km. The concave shape of $F(u)$ is depicted in Figure 1.7.

Let us look at some practical implications of the model. Suppose that at time $y = 0$ there is a traffic jam at some point $x = x_{\mathrm{j}}$. This

could be caused by an accident, a red traffic light, a policeman directing the traffic, etc. Assume further that there is a line

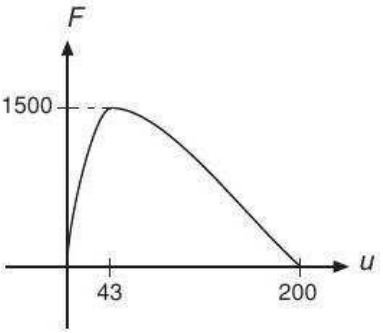

Figure 1.7: The traffic flux F as a function of the density u.

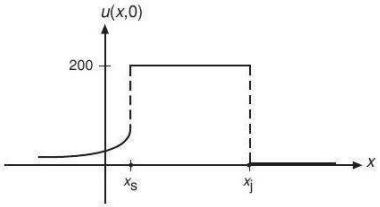

Figure 1.8: The car density at a red traffic light.

of stationary cars extending from x_s to x_j (with $x_j > x_s$). Cars approach the traffic jam from $x < x_s$. At some point the drivers slow down, as they reach the regime where the car density is greater than u_{max}. Therefore the density u just before x_s is as shown schematically in Figure 1.8. The Rankine-Hugoniot condition (1.60) implies that the shock speed is negative. Although the curve $u(x, 0)$ is increasing, the derivative F_u is now negative, and therefore the entropy condition holds. We conclude that a shock wave will propagate from x_s backwards. Indeed

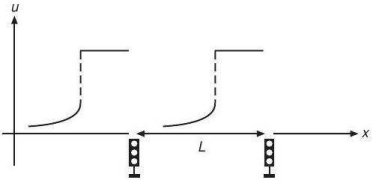

Figure 1.9: Traffic flow through a sequence of traffic lights.

as drivers approach a traffic jam there is a stage when they enter the shock and have to decelerate rapidly. The opposite occurs as we leave the point x_{j}. The density is decreasing and the entropy condition is violated. Therefore we have an expansion wave.

Our analysis can be applied to the design of traffic lights timing. Assume there are several consecutive traffic lights, separated by a distance L (see Figure 1.9). When the traffic approaches one of the traffic lights, a shock propagates backward. To estimate the speed of the shock we assume that behind the shock the density is optimal and so the flux is $F_{\max}$. Then $\gamma_y = -F_{\max}/(u_{\mathrm{b}} - u_{\max})$. Therefore the time it takes the shock to reach the previous traffic light is

$$T_{\mathrm{s}} = \frac{L\,(u_{\mathrm{b}} - u_{\max})}{F_{\max}}.$$

If the red light is maintained over a period exceeding T_s, the high density profile will extend throughout the road, and traffic will come to a complete stop.

1.1.8 The eikonal equation

The two-dimensional eikonal equation takes the form

$$u_x^2 + u_y^2 = n^2, \tag{1.62}$$

where the surfaces $u = c$ (where c is some constant) are the wavefronts, and n is the refraction index of the medium. The initial conditions are given in the form of an initial curve Γ.

To write the characteristic equations, notice that the eikonal equation can be expressed as

$$\left(u_x, u_y, n^2\right) \cdot (u_x, u_y, -1) = 0.$$

Thus the vector $\left(u_x, u_y, n^2\right)$ describes a direction tangent to the solution (integral) surface. To verify this argument algebraically, write equations for the x and y components of the characteristic curve, and check that the equation for the u component is consistent with (1.62).

We thus set

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u_x, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = u_y, \quad \frac{\mathrm{d}u}{\mathrm{d}t} = n^2. \tag{1.63}$$

Since u_x and u_y are unknown at this stage, we compute

$$\begin{aligned} \frac{\mathrm{d}^2x}{\mathrm{d}t^2} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(u_x\right) = u_{xx}\frac{\mathrm{d}x}{\mathrm{d}t} + u_{xy}\frac{\mathrm{d}y}{\mathrm{d}t} = u_{xx}u_x + u_{xy}u_y \\ &= \frac{1}{2}\left(u_x^2 + u_y^2\right)_x = \frac{1}{2}\left(n^2(x,y)\right)_x, \end{aligned} \tag{1.64}$$

and similarly

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = \frac{1}{2}\left(n^2(x,y)\right)_y . \tag{1.65}$$

To write the solution of the eikonal equation, notice that it follows from the definition of the characteristic curves that

$$\frac{\mathrm{d}u}{\mathrm{d}t} = u_x \frac{\mathrm{d}x}{\mathrm{d}t} + u_y \frac{\mathrm{d}y}{\mathrm{d}t} = u_x^2 + u_y^2 = n^2. \tag{1.66}$$

Integrating the last equation leads to a formula that determines u at the point $(x(t), y(t))$ in terms of the initial value of u and the values of the refraction index along the integration path:

$$u(x(t), y(t)) = u(x(0), y(0)) + \int_0^t n^2(x(\tau), y(\tau))\mathrm{d}\tau \tag{1.67}$$

where $(x(t), y(t))$ is a solution of (1.64) and (1.65).

Before solving specific examples, we should clarify an important point regarding the initial conditions for the characteristic equations. Since the original equation (1.63) involves the derivatives of u that are not known at this stage, we eliminated these derivatives by differentiating the characteristic equations once more with respect to the parameter t. Indeed the equations we obtained ((1.64) and (1.65)) no longer depend on u itself; however these are second-order equations! Therefore, it is not enough to provide a single initial condition (such as the initial point of the characteristic curve on the initial curve Γ), but, rather, we must provide the derivatives (x_t, y_t) too. Equivalently, we should provide the vector tangent to the characteristic at the initial point. For this purpose we shall use the fact that the required vector is precisely the gradient

(u_x, u_y) of u. From the eikonal equation itself we know that the size of that vector is $n(x, y)$, and from the initial condition we can find its projection at each point of Γ in the direction tangent to Γ. But obviously the size of a planar vector and its projection along a given direction determine the vector uniquely. Hence we obtain the additional initial condition.

Example 1.1.14. Solve the eikonal equation (1.62) for a medium with a constant refraction index $n = n_0$, and initial condition $u(x, 2x) = 1$. The physical meaning of the initial condition is that the wavefront is a straight line. The characteristic equations are $\mathrm{d}^2x/\mathrm{d}t^2 = 0 = \mathrm{d}^2y/\mathrm{d}t^2$. Thus, the characteristics are straight lines, emanating from the initial line $y = 2x$. Since u is constant on such a line, the gradient of u is orthogonal to it. Hence the second initial condition for the characteristic is

$$\frac{\mathrm{d}x}{\mathrm{d}t}(0) = \frac{2}{\sqrt{5}}n_0, \quad \frac{\mathrm{d}y}{\mathrm{d}t}(0) = -\frac{1}{\sqrt{5}}n_0.$$

We thus obtain:

$$x(t, s) = \frac{2}{\sqrt{5}}n_0 t + x_0(s), \quad y(t, s) = -\frac{1}{\sqrt{5}}n_0 t + y_0(s), \quad u(t, s) = n_0^2 t + u_0(s) \tag{1.68}$$

In order to find $x_0(s)$ and $y_0(s)$, we write the initial curve parameterically as $(s, 2s, 1)$. Substituting the initial curve into (1.68) leads to the integral surface

$$(x, y, u) = \left(\frac{2}{\sqrt{5}}n_0 t + s, -\frac{1}{\sqrt{5}}n_0 t + 2s, n_0^2 t + 1\right). \tag{1.69}$$

Eliminating $t = (2x - y)/\sqrt{5}n_0$, we obtain the explicit solution

$$u(x, y) = 1 + \frac{n_0}{\sqrt{5}}(2x - y)$$

The solution we have obtained has a simple physical interpretation: in a homogeneous medium the characteristic curves are straight lines (classical light rays), and an initial planar wavefront propagates in the direction orthogonal to them. Therefore all wavefronts are planar.

Example 1.1.15. Compute the function $u(x, y)$ satisfying the eikonal equation $u_x^2 + u_y^2 = n^2$ and the initial condition $u(x, 1) = n\sqrt{1 + x^2}$ (n is a constant parameter).

Write the initial conditions parametrically in the form $(x, y, u) = \left(s, 1, n\sqrt{1 + s^2}\right)$. This condition implies $x_t(0, s) = u_x = ns/\sqrt{1 + s^2}$. Substituting the last expression into the eikonal equation gives $y_t(0, s) = u_y = n/\sqrt{1 + s^2}$. Integrating the characteristic equations we obtain

$$(x(t, s), y(t, s), u(t, s)) = \left(\tfrac{ns}{\sqrt{1+s^2}}t + s, \tfrac{n}{\sqrt{1+s^2}}t + 1, n\left(nt + \sqrt{1 + s^2}\ \right)\right).$$

In order to write an explicit solution, observe the identity

$$x^2 + y^2 = \left(nt + \sqrt{1 + s^2}\right)^2$$

satisfied by the integral surface. Therefore, the solution is $u = n\sqrt{x^2 + y^2}$. This solution represents a spherical wave starting from a single point at the origin of coordinates.

1.1.9 General nonlinear equations

The general first-order nonlinear equation takes the form

$$F(x, y, u, u_x, u_y) = 0. \tag{1.70}$$

We shall develop a solution method for such equations. The method is an extension of the method of characteristics. To simplify the presentation we shall use the notation $p = u_x, q = u_y$.

Consider a point (x_0, y_0, u_0) on the integral surface. We want to find the slope of a (characteristic) curve on the integral surface passing through this point. In the quasilinear case the equation determined directly the slope of a specific curve on the integral surface. We shall now construct that curve in a somewhat different manner. Let us write for this purpose the equation of the tangent plane to the integral surface through (x_0, y_0, u_0) :

$$p(x - x_0) + q(y - y_0) - (u - u_0) = 0. \tag{1.71}$$

Notice that the derivatives p and q at (x_0, y_0, u_0) are not independent. Equation (1.70) imposes the relation

$$F(x_0, y_0, u_0, p, q) = 0. \tag{1.72}$$

The last two equations define a one-parameter family of tangent planes. Such a family spreads out a cone. To honor the French geometer Gaspard Monge (1746-1818) this cone is called the *Monge cone*. The natural candidate for the curve we seek in the tangent plane (defining for us the direction of the characteristic curve) is, therefore, the generator of the Monge cone. To compute the generator we differentiate (1.71) by the parameter p :

$$(x - x_0) + \frac{\mathrm{d}q}{\mathrm{d}p}(y - y_0) = 0. \tag{1.73}$$

Assume that F is not degenerate, i.e. $F_p^2 + F_q^2$ never vanishes. Without loss of generality assume $F_q \neq 0$. Then it follows from (1.72) and the implicit function theorem that

$$q'(p) = -\frac{F_p}{F_q} \tag{1.74}$$

Substituting (1.74) into (1.73) we obtain the equation for the Monge cone generator:

$$\frac{x - x_0}{F_p} = \frac{y - y_0}{F_q}. \tag{1.75}$$

Equations (1.71) and (1.75) imply three differential equations for the characteristic curves:

$$\begin{aligned} x_t &= F_p(x, y, u, p, q) \\ y_t &= F_q(x, y, u, p, q) \\ u_t &= pF_p(x, y, u, p, q) + qF_q(x, y, u, p, q) \end{aligned} \tag{1.76}$$

It is easy to verify that equations (1.76) coincide in the quasilinear case with the characteristic equations (1.14). However in the fully nonlinear case the characteristic equations (1.76) do not form a closed system. They contain the hitherto unknown functions p and q. In other words, the characteristic curves carry with them a tangent plane that has to be found as part of the solution. To derive equations for p and q we write

$$p_t = u_{xx}x_t + u_{xy}y_t = u_{xx}F_p + u_{xy}F_q, \tag{1.77}$$

and similarly

$$q_t = u_{yx}F_p + u_{yy}F_q. \tag{1.78}$$

In order to eliminate u_{xx}, u_{xy} and u_{yy}, recall that the equation $F = 0$ holds along the characteristic curves. We therefore obtain upon differentiation

$$F_x + pF_u + p_xF_p + q_xF_q = 0, \tag{1.79}$$

$$F_y + qF_u + p_yF_p + q_yF_q = 0. \tag{1.80}$$

Substituting (1.79) and (1.80) into (1.77) and (1.78) leads to

$$\begin{aligned} p_t &= -F_x - pF_u, \\ q_t &= -F_y - qF_u. \end{aligned}$$

To summarize we write the entire set of characteristic equations:

$$\begin{aligned} x_t &= F_p(x, y, u, p, q), \\ y_t &= F_q(x, y, u, p, q), \\ u_t &= pF_p(x, y, u, p, q) + qF_q(x, y, u, p, q), \\ p_t &= -F_x(x, y, u, p, q) - pF_u(x, y, u, p, q), \\ q_t &= -F_y(x, y, u, p, q) - qF_u(x, y, u, p, q). \end{aligned} \tag{1.81}$$

A simple computation of F_t, using (1.81), indeed verifies that the

PDE holds at all points along a characteristic curve. The main addition to the theory we presented earlier for the quasilinear case is that the characteristic curves have been replaced by more complex geometric structures. Since each characteristic curve now drags with it a tangent plane, we call these structures *characteristic strips*, and equations (1.81) are called the *strip equations*.

We are now ready to formulate the general Cauchy problem for first-order PDEs. Consider (1.70) with the initial condition given by an initial curve $\Gamma \in C^1$:

$$x = x_0(s), \quad y = y_0(s), \quad u = u_0(s). \tag{1.82}$$

We shall show in the proof of the next theorem how to derive initial conditions also for p and q in order to obtain a complete initial value problem for the system (1.81). We do not expect every Cauchy problem to be solvable. Clearly some form of transversality condition must be imposed. It turns out, however, that slightly more than that is required:

Definition 2. Let a point $P_0 = (x_0(s_0), y_0(s_0), u_0(s_0), p_0(s_0), q_0(s_0))$ satisfy the compatibility conditions

$$F(P_0) = 0, \quad u_0'(s_0) = p_0(s_0)\, x_0'(s_0) + q_0(s_0)\, y_0'(s_0) \tag{1.83}$$

If, in addition,

$$x_0'(s_0)\, F_q(P_0) - y_0'(s_0)\, F_p(P_0) \neq 0 \tag{1.84}$$

is satisfied, then we say that the Cauchy problem (1.70), (1.82) satisfies

the *generalized transversality condition* at the point P_0.

Theorem 1.1.16. *Consider the Cauchy problem (1.70), (1.82). Assume that the generalized transversality condition (1.83)-(1.84) holds at P_0. Then there exists $\varepsilon > 0$ and a unique solution $(x(t,s), y(t,s), u(t,s), p(t,s), q(t,s))$ for the Cauchy problem which is defined for $|s - s_o| + |t| < \varepsilon$. Moreover, the parametric representation defines a smooth integral surface $u = u(x,y)$.*

Proof. We start by deriving full initial conditions for the system (1.81). Actually the Cauchy problem has already provided three conditions for (x, y, u) :

$$x(0,s) = x_0(s), \quad y(0,s) = y_0(s), \quad u(0,s) = u_0(s). \tag{1.85}$$

We are left with the task of finding initial conditions

$$p(0,s) = p_0(s), \quad q(0,s) = q_0(s), \tag{1.86}$$

for $p(t,s)$ and $q(t,s)$. Clearly $p_0(s)$ and $q_0(s)$ must satisfy at every point s the differential condition

$$u_0'(s) = p_0(s)x_0'(s) + q_0(s)y_0'(s) \tag{1.87}$$

and the equation itself

$$F\left(x_0(s), y_0(s), u_0(s), p_0(s), q_0(s)\right) = 0. \tag{1.88}$$

However (1.83) guarantees that these two requirements indeed hold at $s = s_0$. The transversality condition (1.84) ensures that the Jacobian of the system (1.87)-(1.88) (with respect to the variables p_0 and q_0) does not vanish at s_0. Therefore, the implicit function theorem implies that one can derive from (1.87)-(1.88) the required initial conditions for p_0 and q_0. Hence the characteristic equations have a full set of initial conditions in a neighborhood $|s - s_0| < \delta$. Since the system of ODEs (1.81) is well-posed, the existence of a unique smooth solution

$$(x(t,s), y(t,s), u(t,s), p(t,s), q(t,s))$$

is guaranteed for (t, s) in a neighborhood of $(0, s_0)$. As in the quasi-linear case, one can verify that (1.81) and (1.88) imply that

$$F(x(t,s), y(t,s), u(t,s), p(t,s), q(t,s)) = 0 \quad \forall\, |s - s_0| < \delta, |t| < \varepsilon. \tag{1.89}$$

In order that the parametric representation for (x, y, u) will define a smooth surface $u(x, y)$, we must show that the mapping $(x(t, s), y(t, s))$ can be inverted to a smooth mapping $(t(x, y), s(x, y))$. Such an inversion exists if the Jacobian $J = \partial(x, y)/\partial(t, s)$ does not vanish. But the characteristic equations imply

$$J|_{(0,s_0)} = \left.\frac{\partial(x,y)}{\partial(s,t)}\right|_{(0,s_0)} = x_s(0,s_0)\, y_t(0,s_0) - x_t(0,s_0)\, y_s(0,s_0) \neq 0, \tag{1.90}$$

where the last inequality follows from the transversality condition

(1.84).

We have thus constructed a smooth function $u(x, y)$. Does it satisfy the Cauchy problem for the nonlinear PDE? This requires that the relations

$$u_t(t, s) = p(t, s)x_t(t, s) + q(t, s)y_t(t, s) \tag{1.91}$$

and

$$u_s(t, s) = p(t, s)x_s(t, s) + q(t, s)y_s(t, s) \tag{1.92}$$

would hold. Condition (1.91) is clearly valid since the characteristic equations were, in fact, constructed to satisfy it. The compatibility condition (1.92) holds on the initial curve, i.e. at $t = 0$. It remains, though, to check this condition also for values of t other than zero. We therefore define the auxiliary function

$$R(t, s) = u_s - px_s - qy_s.$$

We have to show that $R(t, s) = 0$. As we have already argued, the initial data for p and q imply

$$R(0, s) = 0.$$

To check that R also vanishes for other values of t we compute

$$\begin{aligned} R_t &= u_{st} - p_t x_s - p x_{st} - q_t y_s - q y_{st} \\ &= \frac{\partial}{\partial s}\left(u_t - p x_t - q y_t\right) + p_s x_t + q_s y_t - p_t x_s - q_t y_s. \end{aligned}$$

Using (1.91) and the characteristic equations we get

$$\begin{aligned} R_t &= p_s F_p + q_s F_q + x_s\left(F_x + p F_u\right) + y_s\left(F_y + q F_u\right) \\ &= F_u\left(p x_s + q y_s\right) + p_s F_p + q_s F_q + x_s F_x + y_s F_y. \end{aligned}$$

Adding and subtracting $F_u u_s$ to the last expression we find

$$R_t = F_s - F_u R = -F_u R$$

where we used the fact that $F_s = 0$ which follows from (1.89). Since the initial condition for the linear homogeneous ODE $R_t = -F_u R$ is homogeneous too ($R(0, s) = 0$), it follows that $R(t, s) \equiv 0$. □

Example 1.1.17. We write the eikonal equation in the form

$$F(x, y, u, p, q) = p^2 + q^2 - n_0^2 = 0 \tag{1.93}$$

Hence the characteristic equations are

$$\begin{aligned} x_t &= 2p \\ y_t &= 2q, \\ u_t &= 2n_0^2, \\ p_t &= 0, \\ q_t &= 0. \end{aligned} \tag{1.94}$$

The initial conditions for (x, y, u) are given by

$$x(0, s) = s, \quad y(0, s) = 2s, \quad u(0, s) = 1. \tag{1.95}$$

We use (1.87)-(1.88) to derive the initial data for p and q :

$$\begin{aligned} p(0, s) + 2q(0, s) &= 0, \\ p^2(0, s) + q^2(0, s) - n_0^2 &= 0. \end{aligned}$$

Solving these equation we obtain

$$p(0, s) = \frac{2n_0}{\sqrt{5}}, \quad q(0, s) = -\frac{n_0}{\sqrt{5}}. \tag{1.96}$$

It is an easy matter to solve the full characteristic equations:

$$(x, y, u, p, q) = \left(s + \frac{2n_0}{\sqrt{5}}t, 2s - \frac{n_0}{\sqrt{5}}t, 1 + 2n_0^2 t, \frac{2n_0}{\sqrt{5}}, -\frac{n_0}{\sqrt{5}} \right), \tag{1.97}$$

After eliminating t we finally obtain

$$u(x, y) = 1 + \frac{n_0}{\sqrt{5}}(2x - y).$$

Notice that the parametric representation obtained in the current example is different from the one we derived in Example 1.1.14, since the parameter t we used here is half the parameter used in Example 1.1.14.

Exercises

1. Consider the equation $u_x + u_y = 1$, with the initial condition $u(x, 0) = f(x)$.

   (a) What are the projections of the characteristic curves on the (x, y) plane?

   (b) Solve the equation.

2. Solve the equation $xu_x + (x + y)u_y = 1$ with the initial conditions $u(1, y) = y$. Is the solution defined everywhere?

3. Let p be a real number. Consider the PDEs

$$xu_x + yu_y = pu \quad -\infty < x < \infty, \quad -\infty < y < \infty.$$

   (a) Find the characteristic curves for the equations.

   (b) Let $p = 4$. Find an explicit solution that satisfies $u = 1$ on the circle $x^2 + y^2 = 1$.

   (c) Let $p = 2$. Find two solutions that satisfy $u(x, 0) = x^2$, for every $x > 0$.

   (d) Explain why the result in (c) does not contradict the existence-uniqueness theorem.

4. Consider the equation $yu_x - xu_y = 0 (y > 0)$. Check for each of the following initial conditions whether the problem is solvable. If it is solvable, find a solution. If it is not, explain why.

   (a) $u(x, 0) = x^2$.

   (b) $u(x, 0) = x$.

   (c) $u(x, 0) = x, x > 0$.

5. Let $u(x, y)$ be an integral surface of the equation

$$a(x, y)u_x + b(x, y)u_y + u = 0,$$

where $a(x, y)$ and $b(x, y)$ are positive differentiable functions in the entire plane. Define

$$D = \{(x, y), |x| < 1, |y| < 1\}.$$

(a) Prove that the projection on the (x, y) plane of each characteristic curve passing through a point in D intersects the boundary of D at exactly two points.

(b) Show that if u is positive on the boundary of D, then it is positive at every point in D.

(c) Suppose that u attains a local minimum (maximum) at a point $(x_0, y_0) \in D$. Evaluate $u(x_0, y_0)$.

(d) Denote by m the minimal value of u on the boundary of D. Assume $m > 0$. Show that $u(x, y) \geq m$ for all $(x, y) \in D$.

6. The equation $xu_x + \left(x^2 + y\right) u_y + (y/x - x)u = 1$ is given along with the initial condition $u(1, y) = 0$.

(a) Solve the problem for $x > 0$. Compute $u(3, 6)$.

(b) Is the solution defined for the entire ray $x > 0$?

7. Solve the Cauchy problem $u_x + u_y = u^2, u(x, 0) = 1$.

8. (a) Solve the equation $xuu_x + yuu_y = u^2 - 1$ for the ray $x > 0$ under the initial condition $u\left(x, x^2\right) = x^3$.

(b) Is there a unique solution for the Cauchy problem over the entire real line $-\infty < x < \infty$?

9. Consider the equation

$$uu_x + u_y = -\frac{1}{2}u.$$

(a) Show that there is a unique integral surface in a neighborhood of the curve

$$\Gamma_1 = \{(s, 0, \sin s) \mid -\infty < s < \infty\}.$$

(b) Find the parametric representation $x = x(t, s), y = y(t, s), u = u(t, s)$ of the integral surface S for initial condition of part (a).

(c) Find an integral surface S_1 of the same PDE passing through the initial curve

$$\Gamma_1 = \{(s, s, 0) \mid -\infty < s < \infty\}.$$

(d) Find a parametric representation of the intersection curves of the surfaces S and S_1. Hint: Try to characterize that curve relative to the PDE.

10. A river is defined by the domain

$$D = \{(x, y)| \quad |y| < 1, -\infty < x < \infty\}.$$

A factory spills a contaminant into the river. The contaminant is further spread and convected by the flow in the river. The velocity field of the fluid in the river is only in the x direction. The concentration of the contaminant at a point (x, y) in the river and at time τ is denoted by $u(x, y, \tau)$. Conservation of matter and momentum implies that u satisfies the first-order PDE

$$u_\tau - \left(y^2 - 1\right) u_x = 0$$

The initial condition is $u(x, y, 0) = \mathrm{e}^{y}\mathrm{e}^{-x^2}$.

(a) Find the concentration u for all (x, y, τ).

(b) A fish lives near the point $(x, y) = (2, 0)$ at the river. The fish can tolerate contaminant concentration levels up to 0.5. If the concentration exceeds this level, the fish will die at once. Will the fish survive? If yes, explain why. If no, find the time in which the fish will die.

Hint: Notice that y appears in the PDE just as a parameter.

11. Solve the equation $\left(y^2 + u\right) u_x + y u_y = 0$ in the domain $y > 0$, under the initial condition $u = 0$ on the planar curve $x = y^2/2$.

12. Solve the equation $u_y + u^2 u_x = 0$ in the ray $x > 0$ under the initial condition $u(x, 0) = \sqrt{x}$. What is the domain of existence of the solution?

13. Consider the equation $u u_x + x u_y = 1$, with the initial condition $\left(\frac{1}{2}s^2 + 1, \frac{1}{6}s^3 + s, s\right)$.

Find a solution. Are there other solutions? If not, explain why; if there are further solutions, find at least two of them, and explain the lack of uniqueness.

14. Consider the equation $x u_x + y u_y = 1/\cos u$.

(a) Find a solution to the equation that satisfies the condition $u\left(s^2, \sin s\right) = 0$ (you can write down the solution in the implicit form $F(x, y, u) = 0$).

(b) Find some domain of s values for which there exists a unique solution.

15. (a) Find a function $u(x, y)$ that solves the Cauchy problem

$$\left(x + y^2\right) u_x + y u_y + \left(\frac{x}{y} - y\right) u = 1, \quad u(x, 1) = 0 \quad x \in \mathbb{R}.$$

(b) Check whether the transversality condition holds.

(c) Draw the projections on the (x, y) plane of the initial condition and the characteristic curves emanating from the points $(2, 1, 0)$ and $(0, 1, 0)$.

(d) Is the solution you obtained in (a) defined at the origin $(x, y) = (0, 0)$? Explain your answer in light of the existence-uniqueness theorem.

16. Solve the Cauchy problem

$$x u_x + y u_y = -u, \quad u(\cos s, \sin s) = 1 \quad 0 \leq s \leq \pi.$$

Is the solution defined everywhere?

17. Consider the equation

$$x u_x + u_y = 1.$$

(a) Find a characteristic curve passing through the point $(1, 1, 1)$.

(b) Show that there exists a unique integral surface $u(x, y)$ satisfying $u(x, 0) = \sin x$.

(c) Is the solution defined for all x and y ?

18. Consider the equation $uu_x + u_y = -\frac{1}{2}u$.

    (a) Find a solution satisfying $u(x, 2x) = x^2$.

    (b) Is the solution unique?

19. (a) Find a function $u(x, y)$ that solves the Cauchy problem

$$x^2 u_x + y^2 u_y = u^2, \quad u(x, 2x) = x^2 \quad x \in \mathbb{R}.$$

    (b) Check whether the transversality condition holds.

    (c) Draw the projections on the (x, y) plane of the initial curve and the characteristic curves that start at the points $(1, 2, 1)$ and $(0, 0, 0)$.

    (d) Is the solution you found in part (a) defined for all x and y ?

20. Consider the equation

$$yu_x - uu_y = x.$$

    (a) Write a parametric representation of the characteristic curves.

    (b) Solve the Cauchy problem

$$\begin{aligned} yu_x - uu_y &= x, \\ u(s, s) &= -2s \quad -\infty < s < \infty. \end{aligned}$$

    (c) Is the following Cauchy problem solvable:

$$\begin{aligned} yu_x - uu_y &= x, \\ u(s, s) &= s \quad -\infty < s < \infty? \end{aligned}$$

(d) Set

$$w_1 = x + y + u, \quad w_2 = x^2 + y^2 + u^2, \quad w_3 = xy + xu + yu.$$

Show that $w_1 (w_2 - w_3)$ is constant along each characteristic curve.

21. (a) Find a function $u(x, y)$ that solves the Cauchy problem

$$xu_x - yu_y = u + xy, \quad u(x, x) = x^2 \quad 1 \leq x \leq 2.$$

(b) Check whether the transversality condition holds.

(c) Draw the projections on the (x, y) plane of the initial curve and the characteristic curves emanating from the points $(1, 1, 1)$ and $(2, 2, 4)$.

(d) Is the solution you found in (a) well defined in the entire plane?

22. Solve the Cauchy problem $u_x^2 + u_y = 0, u(x, 0) = x$.

23. Let $u(x, t)$ be the solution to the Cauchy problem

$$u_t + cu_x + u^2 = 0, \quad u(x, 0) = x,$$

where c is a constant, t denotes time, and x denotes a space coordinate.

(a) Solve the problem.

(b) A person leaves the point x_0 at time $t = 0$, and moves in the positive x direction with a velocity c (i.e. the quantity $x - ct$

is fixed for him). Show that if $x_0 > 0$, then the solution as seen by the person approaches zero as $t \to \infty$.

(c) What will be observed by such a person if $x_0 < 0$, or if $x_0 = 0$?

24. (a) Solve the problem

$$\begin{aligned} xu_x - uu_y &= y, \\ u(1, y) &= y \qquad -\infty < y < \infty. \end{aligned}$$

(b) Is the solution unique? What is the maximal domain where it is defined?

25. Find at least five solutions for the Cauchy problem

$$u_x + u_y = 1, \quad u(x, x) = x.$$

26. (a) Solve the problem

$$\begin{aligned} xu_y - yu_x + u &= 0, \\ u(x, 0) &= 1 \quad x > 0. \end{aligned}$$

(b) Is the solution unique? What is the maximal domain where it is defined?

27. (a) Use the Lagrange method to find a function $u(x, y)$ that solves the problem

$$\begin{aligned} uu_x + u_y &= 1 \\ u(3x, 0) &= -x \quad -\infty < x < \infty. \end{aligned}$$

(b) Show that the curve $\{(3x, 2, 4 - 3x) \mid -\infty < x < \infty\}$ is contained in the solution surface $u(x, y)$.

(c) Solve

$$\begin{aligned} uu_x + u_y &= 1 \\ u(3x, 2) &= 4 - 3x \quad -\infty < x < \infty \end{aligned}$$

28. Analyze the following problems using the Lagrange method. For each problem determine whether there exists a unique solution, infinitely many solutions or no solution at all. If there is a unique solution, find it; if there are infinitely many solutions, find at least two of them. Present all solutions explicitly.

(a)

$$\begin{aligned} xuu_x + yuu_y &= x^2 + y^2 \quad x > 0, y > 0, \\ u(x, 1) &= \sqrt{x^2 + 1}. \end{aligned}$$

(b)

$$\begin{aligned} xuu_x + yuu_y &= x^2 + y^2 \quad x > 0, \quad y > 0 \\ u(x, x) &= \sqrt{2}x \end{aligned}$$

29. Consider the equation

$$xu_x + (1 + y)u_y = x(1 + y) + xu.$$

(a) Find the general solution.

(b) Assume an initial condition of the form $u(x, 6x - 1) = \phi(x)$. Find a necessary and sufficient condition for ϕ that guarantees the existence of a solution to the problem. Solve the problem for the appropriate ϕ that you found.

(c) Assume an initial condition of the form $u(-1, y) = \psi(y)$. Find a necessary and sufficient condition for ψ that guarantees the existence of a solution to the problem. Solve the problem for the appropriate ψ that you found.

(d) Explain the differences between (b) and (c).

30. (a) Find a compatibility condition for the Cauchy problem

$$u_x^2 + u_y^2 = 1, \quad u(\cos s, \sin s) = 0 \quad 0 \leq s \leq 2\pi.$$

(b) Solve the above Cauchy problem.

(c) Is the solution uniquely defined?

1.2 Second-order linear equations in two independent variables

1.2.1 Introduction

In this section we classify the family of second-order linear equations for functions in two independent variables into three distinct types: hyperbolic (e.g., the wave equation), parabolic (e.g., the heat equation), and elliptic equations (e.g., the Laplace equation). It turns out that solutions of equations of the same type share many exclusive qualitative properties. We show that by a certain change of variables any equation of a particular type can be transformed into a canonical form which is associated with its type.

1.2.2 Classification

We concentrate in this section on second-order linear equations for functions in two independent variables x, y. Such an equation has the form

$$L[u] = au_{xx} + 2bu_{xy} + cu_{yy} + du_x + eu_y + fu = g, \tag{1.98}$$

where $a, b, \ldots, f, g$ are given functions of x, y, and $u(x, y)$ is the unknown function. We introduced the factor 2 in front of the coefficient b for convenience. We assume that the coefficients a, b, c do not vanish simultaneously.

The operator

$$L_0[u] = au_{xx} + 2bu_{xy} + cu_{yy}$$

that consists of the second-(highest-)order terms of the operator L is called the *principal part* of L. It turns out that many fundamental properties of the solutions of (1.98) are determined by its principal part, and, more precisely, by the sign of the discriminant $\delta(L) := b^2 - ac$ of the equation. We classify the equation according to the sign of $\delta(L)$.

Definition 3. Equation (1.98) is said to be *hyperbolic* at a point (x, y) if

$$\delta(L)(x, y) = b(x, y)^2 - a(x, y)c(x, y) > 0,$$

it is said to be *parabolic* at (x, y) if $\delta(L)(x, y) = 0$, and it is said to be *elliptic* at (x, y) if $\delta(L)(x, y) < 0$.

Let Ω be a domain in $\mathbb{R}^2$ (i.e. Ω is an open connected set). The equation is hyperbolic (resp., parabolic, elliptic) in Ω, if it is hyperbolic (resp., parabolic, elliptic) at all points $(x, y) \in \Omega$.

Definition 4. The transformation $(\xi, \eta) = (\xi(x, y), \eta(x, y))$ is called a change of coordinates (or a nonsingular transformation) if the Jacobian $J := \xi_x \eta_y - \xi_y \eta_x$ of the transformation does not vanish at any point (x, y).

Lemma 1.2.1. *The type of a linear second-order PDE in two variables is invariant under a change of coordinates. In other words, the type of the equation is an intrinsic property of the equation and is independent of the particular coordinate system used.*

Proof. Let

$$L[u] = au_{xx} + 2bu_{xy} + cu_{yy} + du_x + eu_y + fu = g, \tag{1.99}$$

and let $(\xi, \eta) = (\xi(x, y), \eta(x, y))$ be a nonsingular transformation. Write $w(\xi, \eta) = u(x(\xi, \eta), y(\xi, \eta))$. We claim that w is a solution of a second-order equation of the same type. Using the chain rule one finds that

$$\begin{aligned}
u_x &= w_\xi \xi_x + w_\eta \eta_x, \\
u_y &= w_\xi \xi_y + w_\eta \eta_y, \\
u_{xx} &= w_{\xi\xi} \xi_x^2 + 2w_{\xi\eta} \xi_x \eta_x + w_{\eta\eta} \eta_x^2 + w_\xi \xi_{xx} + w_\eta \eta_{xx}, \\
u_{xy} &= w_{\xi\xi} \xi_x \xi_y + w_{\xi\eta} (\xi_x \eta_y + \xi_y \eta_x) + w_{\eta\eta} \eta_x \eta_y + w_\xi \xi_{xy} + w_\eta \eta_{xy}, \\
u_{yy} &= w_{\xi\xi} \xi_y^2 + 2w_{\xi\eta} \xi_y \eta_y + w_{\eta\eta} \eta_y^2 + w_\xi \xi_{yy} + w_\eta \eta_{yy}.
\end{aligned}$$

Substituting these formulas into (1.99), we see that w satisfies the following linear equation:

$$\ell[w] := Aw_{\xi\xi} + 2Bw_{\xi\eta} + Cw_{\eta\eta} + Dw_{\xi} + Ew_{\eta} + Fw = G$$

where the coefficients of the principal part of the linear operator ℓ are given by

$$\begin{aligned} A(\xi,\eta) &= a\xi_x^2 + 2b\xi_x\xi_y + c\xi_y^2 \\ B(\xi,\eta) &= a\xi_x\eta_x + b\left(\xi_x\eta_y + \xi_y\eta_x\right) + c\xi_y\eta_y \\ C(\xi,\eta) &= a\eta_x^2 + 2b\eta_x\eta_y + c\eta_y^2 \end{aligned}$$

Notice that we do not need to compute the coefficients of the lower-order derivatives (D, E, F) since the type of the equation is determined only by its principal part (i.e. by the coefficients of the second-order terms). An elementary calculation shows that these coefficients satisfy the following matrix equation:

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} = \begin{pmatrix} \xi_x & \xi_y \\ \eta_x & \eta_y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \xi_x & \eta_x \\ \xi_y & \eta_y \end{pmatrix}.$$

Denote by J the Jacobian of the transformation. Taking the determinant of the two sides of the above matrix equation, we find

$$-\delta(\ell) = AC - B^2 = J^2\left(ac - b^2\right) = -J^2\delta(L).$$

Therefore, the type of the equation is invariant under nonsingular

transformations. □

Three fundamental equations of mathematical physics: the heat equation, the wave equations and the Laplace equation are linear second-order equations. One can easily verify that the **w**ave equation is **h**yperbolic, the **h**eat equation is **p**arabolic, and the **L**aplace equation is **e**lliptic.

We shall show in the next sections that if (1.98) is hyperbolic (resp., parabolic, elliptic) in a domain D, then one can find a coordinate system in which the equation has a simpler form that we call the *canonical form* of the equation. Moreover, in such a case the principal part of the canonical form is equal to the principal part of the fundamental equation of mathematical physics of the same type. This is one of the reasons for studying these fundamental equations.

Definition 5. The canonical form of a hyperbolic equation is

$$\ell[w] = w_{\xi\eta} + \ell_1[w] = G(\xi, \eta),$$

where ℓ_1 is a first-order linear differential operator, and G is a function.

Similarly, the canonical form of a parabolic equation is

$$\ell[w] = w_{\xi\xi} + \ell_1[w] = G(\xi, \eta),$$

and the canonical form of an elliptic equation is

$$\ell[w] = w_{\xi\xi} + w_{\eta\eta} + \ell_1[w] = G(\xi, \eta)$$

1.2.3 Canonical form of hyperbolic equations

Theorem 1.2.1. *Suppose that (1.98) is hyperbolic in a domain D. There exists a coordinate system (ξ, η) in which the equation has the canonical form*

$$w_{\xi\eta} + \ell_1[w] = G(\xi, \eta),$$

where $w(\xi, \eta) = u(x(\xi, \eta), y(\xi, \eta))$, ℓ_1 is a first-order linear differential operator, and G is a function which depends on (1.98).

Proof. Without loss of generality, we may assume that $a(x, y) \neq 0$ for all $(x, y) \in D$. We need to find two functions $\xi = \xi(x, y), \eta = \eta(x, y)$ such that

$$A(\xi, \eta) = a\xi_x^2 + 2b\xi_x\xi_y + c\xi_y^2 = 0,$$
$$C(\xi, \eta) = a\eta_x^2 + 2b\eta_x\eta_y + c\eta_y^2 = 0.$$

The equation that was obtained for the function η is actually the same equation as for ξ; therefore, we need to solve only one equation. It is a first-order equation that is not quasilinear; but as a quadratic form in ξ it is possible to write it as a product of two linear terms

$$\frac{1}{a}\left[a\xi_x + \left(b - \sqrt{b^2 - ac}\right)\xi_y\right]\left[a\xi_x + \left(b + \sqrt{b^2 - ac}\right)\xi_y\right] = 0.$$

Therefore, we need to solve the following linear equations:

$$a\xi_x + \left(b + \sqrt{b^2 - ac}\right)\xi_y = 0 \tag{1.100}$$

$$a\xi_x + \left(b - \sqrt{b^2 - ac}\right)\xi_y = 0. \tag{1.101}$$

In order to obtain a nonsingular transformation $(\xi(x, y), \eta(x, y))$ we choose ξ to be a solution of (1.100) and η to be a solution of (1.101).

These equations are a special case of Example 1.1.4. The characteristic equations for (1.100) are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = b + \sqrt{b^2 - ac}, \quad \frac{\mathrm{d}\xi}{\mathrm{d}t} = 0.$$

Therefore, ξ is constant on each characteristic. The characteristics are solutions of the equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b + \sqrt{b^2 - ac}}{a} \tag{1.102}$$

The function η is constant on the characteristic determined by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b - \sqrt{b^2 - ac}}{a}. \tag{1.103}$$

□

Definition 6. The solutions of (1.102) and (1.103) are called the two families of the *characteristics* (or *characteristic projections*) of the equation $L[u] = g$.

Example 1.2.2. Consider the *Tricomi equation*:

$$u_{xx} + xu_{yy} = 0 \quad x < 0.$$

Find a mapping $q = q(x, y), r = r(x, y)$ that transforms the equation into its canonical form, and present the equation in this coordinate system.

The characteristic equations are

$$\frac{\mathrm{d}y_{\pm}}{\mathrm{d}x} = \pm\sqrt{-x},$$

and their solutions are $\frac{3}{2}y_{\pm} \pm (-x)^{3/2} = \text{constant}$. Thus, the new independent variables are

$$q(x, y) = \frac{3}{2}y + (-x)^{3/2}, \quad r(x, y) = \frac{3}{2}y - (-x)^{3/2}$$

Clearly,

$$q_x = -r_x = -\frac{3}{2}(-x)^{1/2}, \quad q_y = r_y = \frac{3}{2}.$$

Define $v(q, r) = u(x, y)$. By the chain rule

$$\begin{aligned}
u_x &= \frac{-3}{2}(-x)^{1/2}v_q + \frac{3}{2}(-x)^{1/2}v_r, \quad u_y = \frac{3}{2}\left(v_q + v_r\right), \\
u_{xx} &= -\frac{9}{4}xv_{qq} - \frac{9}{4}xv_{rr} + 2\frac{9}{4}xv_{qr} + \frac{3}{4}(-x)^{-1/2}\left(v_q - v_r\right), \\
u_{yy} &= -\frac{9}{4}\left(v_{qq} + 2v_{qr} + v_{rr}\right).
\end{aligned}$$

Substituting these expressions into the Tricomi equation we obtain

$$u_{xx} + xu_{yy} = -9(q-r)^{2/3}\left[v_{qr} + \frac{v_q - v_r}{6(q-r)}\right] = 0.$$

Example 1.2.3. Consider the equation

$$u_{xx} - 2\sin x u_{xy} - \cos^2 x u_{yy} - \cos x u_y = 0. \tag{1.104}$$

Find a coordinate system $s = s(x,y), t = t(x,y)$ that transforms the equation into its canonical form. Show that in this coordinate system the equation has the form $v_{st} = 0$, and find the general solution. The characteristic equations are

$$\frac{\mathrm{d}y_\pm}{\mathrm{d}x} = -\sin x \pm \sqrt{\sin^2 x + \cos^2 x} = -\sin x \pm 1.$$

Consequently, the solutions are $y_\pm = \cos x \pm x +$ constant. The requested transformation is

$$s(x,y) = \cos x + x - y, \quad t(x,y) = \cos x - x - y.$$

Consider now the function $v(s,t) = u(x,y)$ and substitute it into (1.104). We get

$$\begin{aligned}
&\left[v_{ss}(-\sin x + 1)^2 + 2v_{st}(-\sin x + 1)(-\sin x - 1) + v_{tt}(-\sin x - 1)^2\right.\\
&\quad \left. + v_s(-\cos x) + v_t(-\cos x)\right] - 2\sin x\left[v_{ss}(\sin x - 1) + v_{st}(\sin x - 1)\right.\\
&\quad \left. + v_{st}(\sin x + 1) + v_{tt}(\sin x + 1)\right] - \cos^2 x\left[v_{ss} + 2v_{st} + v_{tt}\right]\\
&\quad - \cos x\left(-v_s - v_t\right) = 0.
\end{aligned}$$

Thus, $-4v_{st} = 0$, and the canonical form is

$$v_{st} = 0.$$

It is easily checked that its general solution is $v(s,t) = F(s) + G(t)$, for every $F, G \in C^2(\mathbb{R})$. Therefore, the general solution of (1.104) is

$$u(x,y) = F(\cos x + x - y) + G(\cos x - x - y).$$

1.2.4 Canonical form of parabolic equations

Theorem 1.2.4. *Suppose that (1.98) is parabolic in a domain D. There exists a coordinate system (ξ, η) where the equation has the canonical form*

$$w_{\xi\xi} + \ell_1[w] = G(\xi, \eta),$$

where $w(\xi,\eta) = u(x(\xi,\eta), y(\xi,\eta))$, ℓ_1 is a first-order linear differential operator, and G is a function which depends on (1.98).

Proof. Since $b^2 - ac = 0$, we may assume that $a(x,y) \neq 0$ for all $(x,y) \in D$. We need to find two functions $\xi = \xi(x,y), \eta = \eta(x,y)$ such that $B(\xi,\eta) = C(\xi,\eta) = 0$ for all $(x,y) \in D$. It is enough to make $C = 0$, since the parabolicity of the equation will then imply that $B = 0$. Therefore, we need to find a function η that is a solution of the equation

$$C(\xi,\eta) = a\eta_x^2 + 2b\eta_x\eta_y + c\eta_y^2 = \frac{1}{a}\left(a\eta_x + b\eta_y\right)^2 = 0.$$

From this it follows that η is a solution of the first-order linear equation

$$a\eta_x + b\eta_y = 0. \tag{1.105}$$

Hence, the solution η is constant on each characteristic, i.e., on a curve that is a solution of the equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b}{a}. \tag{1.106}$$

Now, the only constraint on the second independent variable ξ, is that the Jacobian of the transformation should not vanish in D, and we may take any such function ξ. Note that a parabolic equation admits only one family of characteristics while for hyperbolic equations we have two families. □

Example 1.2.5. Prove that the equation

$$x^2u_{xx} - 2xyu_{xy} + y^2u_{yy} + xu_x + yu_y = 0. \tag{1.107}$$

is parabolic and find its canonical form; find the general solution on the half-plane $x > 0$.

We identify $a = x^2, 2b = -2xy, c = y^2$; therefore, $b^2 - ac = x^2y^2 - x^2y^2 = 0$ and the equation is parabolic. The equation for the characteristics is

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

and the solution is $xy =$ constant. Therefore, we define $\eta(x, y) = xy$. The second variable can be simply chosen as $\xi(x, y) = x$. Let $v(\xi, \eta) = u(x, y)$. Substituting the new coordinates ξ and η into (1.107), we obtain

$$x^2\left(y^2 v_{\eta\eta} + 2yv_{\xi\eta} + v_{\xi\xi}\right) - 2xy\left(v_\eta + xyv_{\eta\eta} + xv_{\xi\eta}\right) + x^2 v_{\eta\eta} + xyv_\eta + xv_\xi + xyv_\xi = 0.$$

Thus,

$$\xi^2 v_{\xi\xi} + \xi v_\xi = 0,$$

or $v_{\xi\xi} + (1/\xi)v_\xi = 0$, and this is the desired canonical form.

Setting $w = v_\xi$, we arrive at the first-order ODE $w_\xi + (1/\xi)w = 0$. The solution is $\ln w = -\ln \xi + \tilde{f}(\eta)$, or $w = f(\eta)/\xi$. Hence, v satisfies

$$v = \int v_\xi \mathrm{d}\xi = \int w \ \mathrm{d}\xi = \int \frac{f(\eta)}{\xi}\mathrm{d}\xi = f(\eta)\ln\xi + g(\eta).$$

Therefore, the general solution $u(x, y)$ of (1.107) is $u(x, y) = f(xy)\ln x + g(xy)$, where $f, g \in C^2(\mathbb{R})$ are arbitrary real functions.

1.2.5 Canonical form of elliptic equations

The computation of a canonical coordinate system for the elliptic case is somewhat more subtle than in the hyperbolic case or in the parabolic case. Nevertheless, under the additional assumption that the coefficients of the principal part of the equation are real analytic functions, the procedure for determining the canonical transformation is quite similar to the one for the hyperbolic case.

Definition 7. Let D a planar domain. A function $f : D \to \mathbb{R}$ is said to be real analytic in D if for each point $(x_0, y_0) \in D$, we have a convergent power series expansion

$$f(x,y) = \sum_{k=0}^{\infty} \sum_{j=0}^{k} a_{j,k-j} \, (x - x_0)^j \, (y - y_0)^{k-j} \, ,$$

valid in some neighborhood N of (x_0, y_0).

Theorem 1.2.6. *Suppose that (1.98) is elliptic in a planar domain D. Assume further that the coefficients a, b, c are real analytic functions in D. Then there exists a coordinate system (ξ, η) in which the equation has the canonical form*

$$w_{\xi\xi} + w_{\eta\eta} + \ell_1[w] = G(\xi, \eta),$$

where ℓ_1 is a first-order linear differential operator, and G is a function which depends on (1.98).

Proof. Without loss of generality we may assume that $a(x, y) \neq 0$ for all $(x, y) \in D$. We are looking for two functions $\xi = \xi(x, y), \eta = \eta(x, y)$ that satisfy the equations

$$A(\xi,\eta) = a\xi_x^2 + 2b\xi_x\xi_y + c\xi_y^2 = C(\xi,\eta) = a\eta_x^2 + 2b\eta_x\eta_y + c\eta_y^2, \quad (1.108)$$

$$B(\xi,\eta) = a\xi_x\eta_x + b\,(\xi_x\eta_y + \xi_y\eta_x) + c\xi_y\eta_y = 0. \quad (1.109)$$

This is a system of two nonlinear first-order equations. The main difficulty in the elliptic case is that (1.108)-(1.109) are coupled. In order

to decouple these equations, we shall use the complex plane and the analyticity assumption. We may write the system (1.108)-(1.109) in the following form:

$$a\left(\xi_x^2-\eta_x^2\right)+2b\left(\xi_x\xi_y-\eta_x\eta_y\right)+c\left(\xi_y^2-\eta_y^2\right)=0, \tag{1.110}$$

$$a\xi_x\mathrm{i}\eta_x+b\left(\xi_x\mathrm{i}\eta_y+\xi_y\mathrm{i}\eta_x\right)+c\xi_y\mathrm{i}\eta_y=0, \tag{1.111}$$

where $\mathrm{i}=\sqrt{-1}$. Define the complex function $\phi=\xi+\mathrm{i}\eta$. The system (1.110)-(1.111) is equivalent to the complex valued equation

$$a\phi_x^2+2b\phi_x\phi_y+c\phi_y^2=0.$$

Surprisingly, we have arrived at the same equation as in the hyperbolic case. But in the elliptic case the equation does not admit any real solution, or, in other words, elliptic equations do not have characteristics. As in the hyperbolic case, we factor out the above quadratic PDE, and obtain two linear equations, but now these are complex valued differential equations (where x, y are complex variables!). The nontrivial question of the existence and uniqueness of solutions immediately arises. Fortunately, it is known that if the coefficients of these first-order linear equations are real analytic then it is possible to solve them using the same procedure as in the real case. Moreover, the solutions of the two equations are complex conjugates.

So, we need to solve the equations

$$a\phi_x+\left(b\pm\mathrm{i}\sqrt{ac-b^2}\right)\phi_y=0. \tag{1.112}$$

As before, the solutions ϕ, ψ are constant on the "characteristics" (which are defined on the complex plane):

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b \pm \mathrm{i}\sqrt{ac - b^2}}{a} \tag{1.113}$$

As in the hyperbolic case, the equation in the new coordinates system has the form

$$4v_{\phi\psi} + \cdots = 0.$$

This is still not the elliptic canonical form with real coefficients. We return to our real variables ξ and η using the linear transformation

$$\xi = \operatorname{Re}\phi, \quad \eta = \operatorname{Im}\phi.$$

Since ξ and η are solutions of the system (1.108)-(1.109), it follows that in the variables ξ and η the equation has the canonical form. □

Example 1.2.7. Consider the Tricomi equation:

$$u_{xx} + xu_{yy} = 0, \quad x > 0. \tag{1.114}$$

Find a canonical transformation $q = q(x, y), r = r(x, y)$ and the corresponding canonical form.

The differential equations for the "characteristics" are $\mathrm{d}y/\mathrm{d}x = \pm\sqrt{-x}$, and their solutions are $\frac{3}{2}y \pm \mathrm{i}(x)^{3/2} =$ constant. Therefore, the canonical variables are $q(x, y) = \frac{3}{2}y$ and $r(x, y) = -(x)^{3/2}$. Clearly,

$$q_x = 0, q_y = \frac{3}{2} \quad r_x = -\frac{3}{2}(x)^{1/2}, r_y = 0.$$

Set $v(q, r) = u(x, y)$. Hence,

$$u_x = -\frac{3}{2}(x)^{1/2} v_r, \quad u_y = \frac{3}{2} v_q,$$
$$u_{xx} = \frac{9}{4} x v_{rr} - \frac{3}{4}(x)^{-1/2} v_r, \quad u_{yy} = \frac{9}{4} v_{qq}.$$

Substituting these into the Tricomi equation we obtain the canonical form

$$\frac{1}{x} u_{xx} + u_{yy} = \frac{9}{4}\left(v_{qq} + v_{rr} + \frac{1}{3r} v_r\right) = 0.$$

Exercises

1. Consider the equation

$$u_{xx} - 6u_{xy} + 9u_{yy} = xy^2.$$

   (a) Find a coordinates system (s, t) in which the equation has the form:

$$9v_{tt} = \frac{1}{3}(s - t)t^2.$$

   (b) Find the general solution $u(x, y)$.

   (c) Find a solution of the equation which satisfies the initial conditions $u(x, 0) = \sin x, u_y(x, 0) = \cos x$ for all $x \in \mathbb{R}$.

2. (a) Show that the following equation is hyperbolic:

$$u_{xx} + 6u_{xy} - 16u_{yy} = 0.$$

(b) Find the canonical form of the equation.

(c) Find the general solution $u(x, y)$.

(d) Find a solution $u(x, y)$ that satisfies $u(-x, 2x) = x$ and $u(x, 0) = \sin 2x$.

3. Consider the equation

$$u_{xx} + 4u_{xy} + u_x = 0.$$

(a) Bring the equation to a canonical form.

(b) Find the general solution $u(x, y)$ and check by substituting back into the equation that your solution is indeed correct.

(c) Find a specific solution satisfying

$$u(x, 8x) = 0, \quad u_x(x, 8x) = 4\mathrm{e}^{-2x}.$$

4. Consider the equation

$$y^5 u_{xx} - y u_{yy} + 2u_y = 0, \quad y > 0.$$

(a) Find the canonical form of the equation.

(b) Find the general solution $u(x, y)$ of the equation.

(c) Find the solution $u(x, y)$ which satisfies $u(0, y) = 8y^3$, and $u_x(0, y) = 6$, for all $y > 0$.

5. Consider the equation

$$xu_{xx} - yu_{yy} + \frac{1}{2}(u_x - u_y) = 0.$$

(a) Find the domain where the equation is elliptic, and the domain where it is hyperbolic

(b) For each of the above two domains, find the corresponding canonical transformation.

6. Consider the equation

$$u_{xx} + \left(1 + y^2\right)^2 u_{yy} - 2y\left(1 + y^2\right) u_y = 0.$$

(a) Find the canonical form of the equation.

(b) Find the general solution $u(x, y)$ of the equation.

(c) Find the solution $u(x, y)$ which satisfies $u(x, 0) = g(x)$, and $u_y(x, 0) = f(x)$, where $f, g \in C^2(\mathbb{R})$.

(d) Find the solution $u(x, y)$ for $f(x) = -2x$, and $g(x) = x$.

7. Consider the equation

$$u_{xx} + 2u_{xy} + [1 - q(y)]u_{yy} = 0$$

where

$$q(y) = \begin{cases} -1 & y < -1 \\ 0 & |y| \leq 1 \\ 1 & y > 1 \end{cases}$$

(a) Find the domains where the equation is hyperbolic, parabolic, and elliptic.

(b) For each of the above three domains, find the corresponding canonical transformation and the canonical form.

(c) Draw the characteristics for the hyperbolic case.

8. Consider the equation

$$4y^2u_{xx} + 2\left(1 - y^2\right)u_{xy} - u_{yy} - \frac{2y}{1+y^2}\left(2u_x - u_y\right) = 0.$$

(a) Find the canonical form of the equation.

(b) Find the general solution $u(x, y)$ of the equation.

(c) Find the solution $u(x, y)$ which satisfies $u(x, 0) = g(x)$, and $u_y(x, 0) = f(x)$, where $f, g \in C^2(\mathbb{R})$ are arbitrary functions.

9. (a) Prove that in the hyperbolic case the canonical transformation is nonsingular ($J \neq 0$).

(b) Prove that in the elliptic case the canonical transformation is nonsingular ($J \neq 0$).

10. Consider the equation

$$u_{xx} - 2u_{xy} + 4\mathrm{e}^y = 0.$$

(a) Find the canonical form of the equation.

(b) Find the solution $u(x, y)$ which satisfies $u(0, y) = f(y)$, and $u_x(0, y) = g(y)$.

11. In continuation of Example 1.2.3, consider the equation

$$u_{xx} - 2\sin x u_{xy} - \cos^2 x u_{yy} - \cos x u_y = 0.$$

(a) Find a solution of the equation which satisfies $u(0, y) = f(y)$, $u_x(0, y) = g(y)$, where f, g are given functions.

(b) Find conditions on f and g such that the solution $u(x, y)$ of part (a) is a classical solution.

12. Consider the equation

$$u_{xx} + yu_{yy} = 0$$

Find the canonical forms of the equation for the domain where the equation is hyperbolic, and for the domain where it is elliptic.

1.3 The one-dimensional wave equation

1.3.1 Introduction

In this section we study the one-dimensional wave equation on the real line. The canonical form of the wave equation will be used to show that the Cauchy problem is well-posed. Moreover, we shall derive simple explicit formulas for the solutions. We also discuss some important properties of the solutions of the wave equation which are typical for more general hyperbolic problems as well.

1.3.2 Canonical form and general solution

The homogeneous wave equation in one (spatial) dimension has the form

$$u_{tt} - c^2 u_{xx} = 0 \quad -\infty \leq a < x < b \leq \infty, t > 0, \tag{1.115}$$

where $c \in \mathbb{R}$ is called the *wave speed*, a terminology that will be justified in the discussion below.

To obtain the canonical form of the wave equation, define the new variables

$$\xi = x + ct \quad \eta = x - ct,$$

and set $w(\xi, \eta) = u(x(\xi, \eta), t(\xi, \eta))$. Using the chain rule for the function $u(x, t) = w(\xi(x, t), \eta(x, t))$, we obtain

$$u_t = w_\xi \xi_t + w_\eta \eta_t = c\left(w_\xi - w_\eta\right), \quad u_x = w_\xi \xi_x + w_\eta \eta_x = w_\xi + w_\eta,$$

and

$$u_{tt} = c^2\left(w_{\xi\xi} - 2w_{\xi\eta} + w_{\eta\eta}\right), \quad u_{xx} = w_{\xi\xi} + 2w_{\xi\eta} + w_{\eta\eta}.$$

Hence,

$$u_{tt} - c^2 u_{xx} = -4c^2 w_{\xi\eta} = 0.$$

This is the canonical form for the wave equation. Since $\left(w_\xi\right)_\eta = 0$, it follows that $w_\xi = f(\xi)$, and then $w = \int f(\xi)\mathrm{d}\xi + G(\eta)$. Therefore, the general solution of the equation $w_{\xi\eta} = 0$ has the form

$$w(\xi, \eta) = F(\xi) + G(\eta),$$

where $F, G \in C^2(\mathbb{R})$ are two arbitrary functions. Thus, in the original

variables, the general solution of the wave equation is

$$u(x,t) = F(x+ct) + G(x-ct). \tag{1.116}$$

In other words, if u is a solution of the one-dimensional wave equation, then there exist two real functions $F, G \in C^2$ such that (1.116) holds. Conversely, any two functions $F, G \in C^2$ define a solution of the wave equation via formula (1.116).

For a fixed $t_0 > 0$, the graph of the function $G(x - ct_0)$ has the same shape as the graph of the function $G(x)$, except that it is shifted to the right by a distance ct_0. Therefore, the function $G(x - ct)$ represents a wave moving to the right with velocity c, and it is called a *forward wave*. The function $F(x + ct)$ is a wave traveling to the left with the same speed, and it is called a *backward wave*. Indeed c can be called the *wave speed*.

Equation (1.116) demonstrates that any solution of the wave equation is the sum of two such traveling waves. This observation will enable us to obtain graphical representations of the solutions (the graphical method).

We would like to extend the validity of (1.116). Observe that for any two real piecewise continuous functions F, G, (1.116) defines a piecewise continuous function u that is a superposition of a forward wave and a backward wave traveling in opposite directions with speed c. Moreover, it is possible to find two sequences of smooth functions, $\{F_n(s)\}, \{G_n(s)\}$, converging at any point to F and G, respectively, which converge uniformly to these functions in any bounded and closed interval that does not contain points of discontinuity. The function

$$u_n(x,t) = F_n(x+ct) + G_n(x-ct)$$

is a proper solution of the wave equation, but the limiting function $u(x,t) = F(x+ ct) + G(x - ct)$ is not necessarily twice differentiable, and therefore might not be a solution. We call a function $u(x,t)$ that satisfies (1.116) with piecewise continuous functions F, G a *generalized solution* of the wave equation.

Let us further discuss the general solution (1.116). Consider the (x,t) plane. The following two families of lines

$$x - ct = \text{ constant}, \quad x + ct = \text{ constant},$$

are called the characteristics of the wave equation. For the wave equation, the characteristics are straight lines in the (x,t) plane with slopes $\pm 1/c$. It turns out that as for first-order PDEs, the "information" is transferred via these curves.

We arrive now at one of the most important properties of the characteristics. Assume that for a fixed time t_0, the solution u is a smooth function except at one point (x_0, t_0). Clearly, either F is not smooth at $x_0 + ct_0$, and/or the function G is not smooth at $x_0 - ct_0$. There are two characteristics that pass through the point (x_0, t_0); these are the lines

$$x - ct = x_0 - ct_0, \quad x + ct = x_0 + ct_0.$$

Consequently, for any time $t_1 \neq t_0$ the solution u is smooth except at one or two points $x_\pm$ that satisfy

$$x_- - ct_1 = x_0 - ct_0, \quad x_+ + ct_1 = x_0 + ct_0.$$

Therefore, the singularities (nonsmoothness) of solutions of the wave equation are traveling only along characteristics. This phenomenon is typical of hyperbolic equations in general: a singularity is not smoothed out; rather it travels at a finite speed. This is in contrast to parabolic and elliptic equations, where, as will be shown in the following chapters, singularities are immediately smoothed out.

Example 1.3.1. Let $u(x,t)$ be a solution of the wave equation

$$u_{tt} - c^2 u_{xx} = 0,$$

which is defined in the whole plane. Assume that u is constant on the line $x = 2 + ct$. Prove that $u_t + cu_x = 0$.

The solution $u(x,t)$ has the form $u(x,t) = F(x+ct) + G(x-ct)$. Since $u(2+ ct, t) =$ constant, it follows that

$$F(2 + 2ct) + G(2) = \text{ constant } .$$

Set $s = 2 + 2ct$, we have $F(s) =$ constant. Consequently $u(x,t) = G(x - ct)$. Computing now the expression $u_t + cu_x$, we obtain

$$u_t + cu_x = -cG'(x - ct) + cG'(x - ct) = 0.$$

1.3.3 The Cauchy problem and d'Alembert's formula

The Cauchy problem for the one-dimensional homogeneous wave equation is given by

$$u_{tt} - c^2 u_{xx} = 0 \quad -\infty < x < \infty, t > 0, \tag{1.117}$$

$$u(x,0) = f(x), \quad u_t(x,0) = g(x), \quad -\infty < x < \infty. \tag{1.118}$$

A solution of this problem can be interpreted as the amplitude of a sound wave propagating in a very long and narrow pipe, which in practice can be considered as a one-dimensional infinite medium. This system also represents the vibration of an infinite (ideal) string. The initial conditions f, g are given functions that represent the amplitude u, and the velocity u_t of the string at time $t = 0$.

A classical (proper) solution of the Cauchy problem (1.117)-(1.118) is a function u that is continuously twice differentiable for all $t > 0$, such that u and u_t are continuous in the half-space $t \geq 0$, and such that (1.117)-(1.118) are satisfied. Generally speaking, classical solutions should have the minimal smoothness properties in order to satisfy continuously all the given conditions in the classical sense.

Recall that the general solution of the wave equation is of the form

$$u(x,t) = F(x + ct) + G(x - ct). \tag{1.119}$$

Our aim is to find F and G such that the initial conditions of (1.118) are satisfied. Substituting $t = 0$ into (1.119) we obtain

$$u(x,0) = F(x) + G(x) = f(x). \tag{1.120}$$

Differentiating (1.119) with respect to t and substituting $t = 0$, we have

$$u_t(x,0) = cF'(x) - cG'(x) = g(x). \tag{1.121}$$

Integration of (1.121) over the integral $[0, x]$ yields

$$F(x) - G(x) = \frac{1}{c}\int_0^x g(s)\mathrm{d}s + C, \tag{1.122}$$

where $C = F(0) - G(0)$. Equations (1.120) and (1.122) are two linear algebraic equations for $F(x)$ and $G(x)$. The solution of this system of equations is given by

$$F(x) = \frac{1}{2}f(x) + \frac{1}{2c}\int_0^x g(s)\mathrm{d}s + \frac{C}{2}, \tag{1.123}$$

$$G(x) = \frac{1}{2}f(x) - \frac{1}{2c}\int_0^x g(s)\mathrm{d}s - \frac{C}{2}. \tag{1.124}$$

By substituting these expressions for F and G into the general solution (1.119), we obtain the formula

$$u(x,t) = \frac{f(x+ct) + f(x-ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)\mathrm{d}s, \tag{1.125}$$

which is called *d'Alembert's formula*. Note that sometimes (1.123)-

(1.124) are also useful, as they give us explicit formulas for the forward and the backward waves.

The following examples illustrate the use of d'Alembert's formula.

Example 1.3.2. Consider the Cauchy problem

$$u_{tt} - u_{xx} = 0 \quad -\infty < x < \infty, \quad t > 0,$$

$$u(x,0) = f(x) = \begin{cases} 0 & -\infty < x < -1, \\ x+1 & -1 \le x \le 0, \\ 1-x & 0 \le x \le 1, \\ 0 & 1 < x < \infty, \end{cases}$$

$$u_t(x,0) = g(x) = \begin{cases} 0 & -\infty < x < -1, \\ 1 & -1 \le x \le 1, \\ 0 & 1 < x < \infty. \end{cases}$$

(a) Evaluate u at the point $\left(1, \frac{1}{2}\right)$.

(b) Discuss the smoothness of the solution u.

(a) Using d'Alembert's formula, we find that

$$u\left(1, \frac{1}{2}\right) = \frac{f\left(\frac{3}{2}\right) + f\left(\frac{1}{2}\right)}{2} + \frac{1}{2}\int_{\frac{1}{2}}^{\frac{3}{2}} g(s)\mathrm{d}s.$$

Since $\frac{3}{2} > 1$ it follows that $f\left(\frac{3}{2}\right) = 0$. On the other hand, $0 \le \frac{1}{2} \le 1$; therefore, $f\left(\frac{1}{2}\right) = \frac{1}{2}$. Evidently, $\int_{\frac{1}{2}}^{\frac{3}{2}} g(s)\mathrm{d}s = \int_{\frac{1}{2}}^{1} 1 \, \mathrm{d}s = \frac{1}{2}$. Thus, $u\left(1, \frac{1}{2}\right) = \frac{1}{2}$.

(b) The solution is not classical, since $u \notin C^1$. Yet u is a generalized solution of the problem. Note that although g is not continuous, nevertheless the solution u is a continuous function. The singularities of the solution propagate along characteristics that intersect the initial line $t = 0$ at the singularities of the initial conditions. These are exactly the characteristics $x \pm t = -1, 0, 1$. Therefore, the solution is smooth in a neighborhood of the point $\left(1, \frac{1}{2}\right)$ which does not intersect these characteristics.

Example 1.3.3. Let $u(x,t)$ be the solution of the Cauchy problem

$$u_{tt} - 9u_{xx} = 0 \quad -\infty < x < \infty, t > 0,$$
$$u(x,0) = f(x) = \begin{cases} 1 & |x| \leq 2, \\ 0 & |x| > 2, \end{cases}$$
$$u_t(x,0) = g(x) = \begin{cases} 1 & |x| \leq 2, \\ 0 & |x| > 2. \end{cases}$$

(a) Find $u\left(0, \frac{1}{6}\right)$.

(b) Discuss the large time behavior of the solution.

(c) Find the maximal value of $u(x,t)$, and the points where this maximum is achieved.

(d) Find all the points where $u \in C^2$.

(a) Since

$$u(x,t) = \frac{f(x+3t) + f(x-3t)}{2} + \frac{1}{6}\int_{x-3t}^{x+3t} g(s)\mathrm{d}s.$$

it follows that for $x = 0$ and $t = \frac{1}{6}$, we have

$$u\left(0, \frac{1}{6}\right) = \frac{f\left(\frac{1}{2}\right) + f\left(-\frac{1}{2}\right)}{2} + \frac{1}{6}\int_{-\frac{1}{2}}^{\frac{1}{2}} g(s)\mathrm{d}s$$
$$= \frac{1+1}{2} + \frac{1}{6}\int_{-\frac{1}{2}}^{\frac{1}{2}} 1 \ \mathrm{d}s = \frac{7}{6}.$$

(b) Fix $\xi \in \mathbb{R}$ and compute $\lim_{t\to\infty} u(\xi, t)$. Clearly,

$$\lim_{t\to\infty} f(\xi + 3t) = 0, \quad \lim_{t\to\infty} f(\xi - 3t) = 0,$$
$$\lim_{t\to\infty} \int_{\xi-3t}^{\xi+3t} g(s)\mathrm{d}s = \int_{-2}^{2} 1 \ \mathrm{d}s = 4.$$

Therefore, $\lim_{t\to\infty} u(\xi, t) = \frac{2}{3}$.

(c) Recall that for any real functions f, g,

$$\max\{f(x) + g(x)\} \leq \max f(x) + \max g(x).$$

It turns out that in our special case there exists a point (x, t), where all the terms in (1.125) attain their maximal value simultaneously, and therefore at such a point the maximum of u is attained.

Indeed, $\max\{f(x + 3t)\} = 1$ which is attained on the strip $-2 \leq x + 3t \leq 2$. Similarly, $\max\{f(x - 3t)\} = 1$ which is attained on the strip $-2 \leq x - 3t \leq 2$, while $\max\left\{\int_{x-3t}^{x+3t} g(s)\mathrm{d}s = \int_{-2}^{2} 1ds = 4\right.$, and it is attained on the intersection of the halfplanes $x + 3t \geq 2$ and $x - 3t \leq -2$. The intersection of all these sets is the set of all points that satisfy the two equations

$$x + 3t = 2,$$
$$x - 3t = -2.$$

This system has a unique solution at $(x, t) = \left(0, \frac{2}{3}\right)$. Thus, the solution u achieves its maximum at the point $\left(0, \frac{2}{3}\right)$, where $u\left(0, \frac{2}{3}\right) = \frac{5}{3}$.

(d) The initial conditions are smooth except at the points $x = \pm 2$. Therefore, the solution is smooth at all points that are not on the characteristics

$$x \pm 3t = -2, \quad x \pm 3t = 2.$$

The function u is a generalized solution that is piecewise continuous for any fixed time $t > 0$.

The well-posedness of the Cauchy problem follows from the d'Alembert formula.

Theorem 1.3.4. *Fix $T > 0$. The Cauchy problem (1.117)-(1.118) in the domain $-\infty < x < \infty, 0 \le t \le T$ is well-posed for $f \in C^2(\mathbb{R}), g \in C^1(\mathbb{R})$.*

Proof. The existence and uniqueness follow directly from the d'Alembert formula. Indeed, this formula provides us with a solution, and we have shown that any solution of the Cauchy problem is necessarily equal to the d'Alembert solution. Note that from our smoothness assumption $\left(f \in C^2(\mathbb{R}), g \in C^1(\mathbb{R})\right)$, it follows that $u \in C^2(\mathbb{R} \times (0, \infty)) \cap C^1(\mathbb{R} \times [0, \infty))$, and therefore, the d'Alembert solution is a classical solution. On the other hand, for $f \in C(\mathbb{R})$ and g that is locally integrable, the d'Alembert solution is a generalized solution.

It remains to prove the stability of the Cauchy problem, i.e. we need to show that small changes in the initial conditions give rise to a small change in the solution. Let u_i be two solutions of the Cauchy problem with initial conditions f_i, g_i, where $i = 1, 2$. Now, if

$$|f_1(x) - f_2(x)| < \delta, \quad |g_1(x) - g_2(x)| < \delta,$$

for all $x \in \mathbb{R}$, then for all $x \in \mathbb{R}$ and $0 \leq t \leq T$ we have

$$|u_1(x,t) - u_2(x,t)| \leq \frac{|f_1(x+ct) - f_2(x+ct)|}{2} + \frac{|f_1(x-ct) - f_2(x-ct)|}{2}$$
$$+ \frac{1}{2c}\int_{x-ct}^{x+ct} |g_1(s) - g_2(s)|\,\mathrm{d}s < \frac{1}{2}(\delta + \delta) + \frac{1}{2c}2ct\delta \leq (1+T)\delta.$$

Therefore, for a given $\varepsilon > 0$, we take $\delta < \varepsilon/(1+T)$. Then for all $x \in \mathbb{R}$ and $0 \leq t \leq T$ we have

$$|u_1(x,t) - u_2(x,t)| < \varepsilon.$$

□

Remark 1.3.1.

(1) The Cauchy problem is ill-posed on the domain $-\infty < x < \infty, t \geq 0$.

(2) The d'Alembert formula is also valid for $-\infty < x < \infty, T < t \leq 0$, and the Cauchy problem is also well-posed in this domain. The physical interpretation is that the process is reversible.

1.3.4 Domain of dependence and region of influence

Let us return to the Cauchy problem (1.117)-(1.118), and examine what is the information that actually determines the solution u at a fixed point (x_0, t_0). Consider the (x, t) plane and the two characteristics passing through the point (x_0, t_0) :

$$x - ct = x_0 - ct_0, \quad x + ct = x_0 + ct_0.$$

These straight lines intersect the x axis at the points $(x_0 - ct_0, 0)$ and $(x_0 + ct_0, 0)$, respectively. The triangle formed by the these characteristics and the interval $[x_0 - ct_0, x_0 + ct_0]$ is called a *characteristic triangle* (see Figure 1.10).

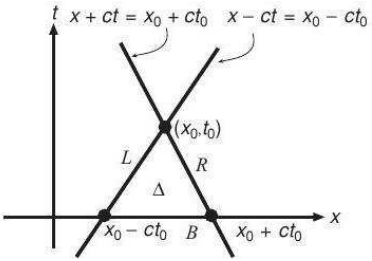

Figure 1.10: Domain of dependence.

By the d'Alembert formula

$$u(x_0, t_0) = \frac{f(x_0 + ct_0) + f(x_0 - ct_0)}{2} + \frac{1}{2c} \int_{x_0 - ct_0}^{x_0 + ct_0} g(s) \mathrm{d}s. \quad (1.126)$$

Therefore, the value of u at the point (x_0, t_0) is determined by the

values of f at the vertices of the characteristic base and by the values of g along this base. Thus, $u(x_0, t_0)$ depends only on the part of the initial data that is given on the interval $[x_0 - ct_0, x_0 + ct_0]$. Therefore, this interval is called *domain of dependence* of u at the point (x_0, t_0). If we change the initial data at points outside this interval, the value of the solution u at the point (x_0, t_0) will not change. Information on a change in the data travels with speed c along the characteristics, and therefore such information is not available for $t \leq t_0$ at the point x_0. The change will finally influence the solution at the point x_0 at a later time. Hence, for every point (x, t) in a fixed characteristic triangle, $u(x, t)$ is determined only by the initial data that are given on (part of) the characteristic base (see Figure 1.10). Furthermore, if the initial data are smooth on this base, then the solution is smooth in the whole triangle.

We may ask now the opposite question: which are the points on the half-plane $t > 0$ that are influenced by the initial data on a fixed interval $[a, b]$? The set of all such points is called the *region of influence* of the interval $[a, b]$. It follows from the discussion above that the points of this interval influence the value of the solution u at a point (x_0, t_0) if and only if $[x_0 - ct_0, x_0 + ct_0] \cap [a, b] \neq \emptyset$. Hence the initial data along the interval $[a, b]$ influence only points (x, t) satisfying

$$x - ct \leq b, \text{ and } x + ct \geq a.$$

These are the points inside the forward (truncated) characteristic cone that is defined by the base $[a, b]$ and the edges $x + ct = a, x - ct = b$ (it is the union of the regions I-IV of Figure 1.11).

Assume, for instance, that the initial data f, g vanish outside the interval $[a, b]$. Then the amplitude of the vibrating string is zero at

every point outside the influence

region of this interval. On the other hand, for a fixed point x_0 on the string, the effect of the perturbation (from the zero data) along the interval $[a, b]$ will be felt after a time $t_0 \geq 0$, and eventually, for t large enough, the solution takes the constant value $u(x_0, t) = (1/2c) \int_a^b g(s)\mathrm{d}s$. This occurs precisely at points (x_0, t) that are inside the cone

$$x_0 - ct \leq a, \text{ and } x_0 + ct \geq b,$$

(see region IV in Figure 1.11). Using these observations, we demonstrate in the following example the so-called graphical method for solving the Cauchy problem for the wave equation.

Example 1.3.5. Consider the Cauchy problem

$$u_{tt} - c^2 u_{xx} = 0 \quad -\infty < x < \infty, t > 0,$$
$$u(x,0) = f(x) = \begin{cases} 2 & |x| \leq a, \\ 0 & |x| > a, \end{cases}$$
$$u_t(x,0) = g(x) = 0 \quad -\infty < x < \infty.$$

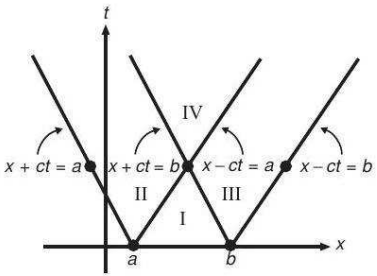

Figure 1.11: Region of influence.

Draw the graphs of the solution $u(x,t)$ at times $t_i = ia/2c$, where $i = 0, 1, 2, 3$. Using d'Alembert's formula, we write the solution u as a sum of backward and forward waves

$$u(x,t) = \frac{f(x+ct) + f(x-ct)}{2}.$$

Since these waves are piecewise constant functions, it is clear that for each t, the solution u is also a piecewise constant function of x with values $u = 0$, 1,2 .

Consider the (x,t) plane. We draw the characteristic lines that pass through the special points on the initial line $t = 0$ where the initial data are not smooth. In the present problem these are the points $x = \pm a$. We also draw the lines $t = t_i$ that will serve us as the abscissas (x axes) for the graphs of the functions $u\,(x,t_i)$. Note that the ordinate of the coordinate system is used as the t and the u axes (see Figure 1.12).

Consider the time $t = t_1$. The forward wave has traveled $a/2$ units to the right, and the backward wave has traveled $a/2$ units to the left. The support (the set of points where the function is not zero) of the forward wave at time t_1 is the interval $[-a/2, 3a/2]$, while $[-3a/2, a/2]$ is the support of the backward wave at t_1. Therefore, the support of the solution at t_1 is the interval $[-3a/2, 3a/2]$, i.e. the region of influence of $[-a, a]$ at $t = t_1$. Now, at the intersection of the supports of the two waves (the interval $[-a/2, a/2]$) u takes the value $1+1 = 2$, while on the intervals $[-3a/2, -a/2), (a/2, 3a/2]$, where the supports do not intersect, u takes the value 1 . Obviously, $u = 0$ at all other points.

Consider the time $t_2 = a/c$. The support of the forward (backward) wave is $[0, 2a]$ ($[-2a, 0]$, respectively). Consequently, the support of the solution u is $[-2a, 2a]$, i.e. the region of influence of $[-a, a]$ at $t = t_2$.

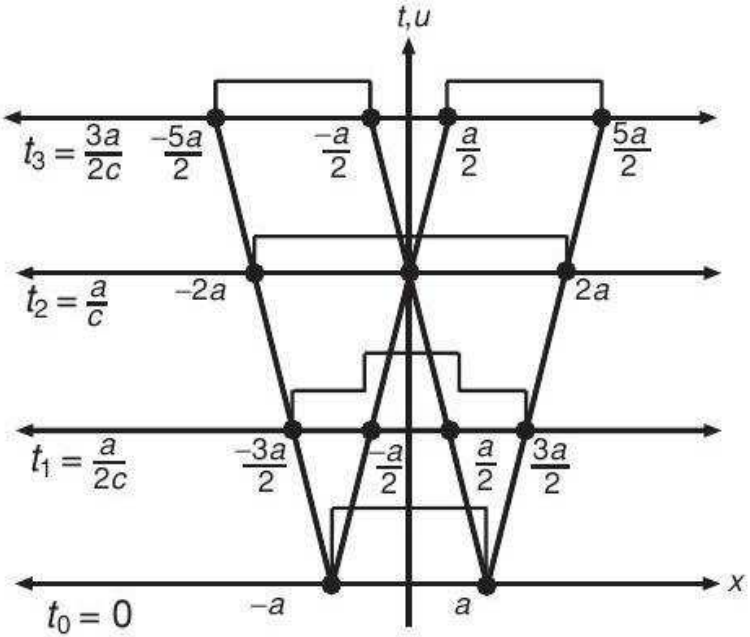

Figure 1.12: The graphical method.

The intersection of the supports of the two waves is the point $x = 0$, where u takes the value 2 . On the intervals $[-2a, 0), (0, 2a], u$ is 1 . Obviously, $u = 0$ at all other points.

At the time $t_3 = 3a/2c$, the support of the forward (backward) wave is $[a/2, 5a/2]$ ($[-5a/2, -a/2]$, respectively), and there is no interaction between the waves. Therefore, the solution at these intervals equals 1 , and it equals zero otherwise.

To conclude, the first step of the graphical method is to compute and to draw the graphs of the forward and backward waves. Then, for a given time t, we shift these two shapes to the right and, respectively,

to the left by ct units. Finally, we add the two graphs.

In the next example we use the graphical method to investigate the influence of the initial velocity on the solution.

Example 1.3.6. Find the graphs of the solution $u(x,t_i)\,, \quad t_i = i, i = 1,4$ for the problem

$$u_{tt} - u_{xx} = 0 \quad -\infty < x < \infty, t > 0,$$

$$u(x,0) = f(x) = 0, \quad -\infty < x < \infty,$$

$$u_t(x,0) = g(x) = \begin{cases} 0 & x < 0, \\ 1 & x \geq 0. \end{cases}$$

We apply d'Alembert's formula to write the solution as the sum of forward and backward waves:

$$u(x,t) = \frac{1}{2}\int_{x-t}^{0} g(s)\mathrm{d}s + \frac{1}{2}\int_{0}^{x+t} g(s)\mathrm{d}s = -\frac{\max\{0, x-t\}}{2} + \frac{\max\{0, x+t\}}{2}.$$

Since both the forward and backward waves are piecewise linear functions, the solution $u(x,t)$ for all times t is a piecewise linear function of x.

We draw in the plane (x,t) the characteristics emanating from the points where the initial condition is nonsmooth. In our case this happens at just one point, namely $x = 0$. We also depict the lines $t = t_i$ that form the abscissas for the graph of $u(x,t_i)$ (see Figure 1.13).

At time t_1, the forward wave has moved one unit to the right, and the backward wave has moved one unit to the left. The forward wave is

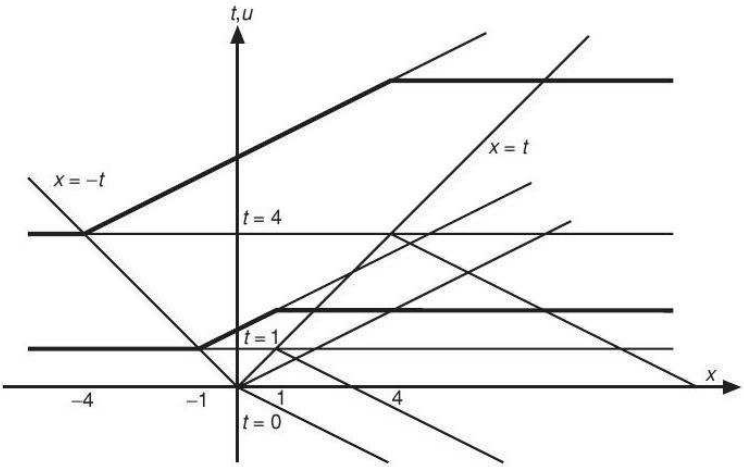

Figure 1.13: The graphical solution for Example 1.3.6

.

supported on the interval $[1, \infty)$, while the backward wave is supported on $[-1, \infty)$. Therefore the solution is supported on $[-1, \infty)$. It is clear that on the interval $[-1, 1]$ the solution forms a linear function thanks to the backward wave. Specifically, $u = (x + 1)/2$ there. In the interval $[1, \infty)$ the forward wave is a linear function with slope $-1/2$, while the backward wave is a linear function with slope $1/2$. Therefore the solution itself, being a superposition of these two waves, is constant $u = 1$ there.

A similar consideration can be used to draw the graph for $t = 4$. Actually, the solution can be written explicitly as

$$u(x,t) = \begin{cases} 0 & x < -t, \\ \frac{x+t}{2} & -t \le x \le t, \\ t & x > t. \end{cases}$$

Notice that for each fixed $x_0 \in \mathbb{R}$, the solution $u(x_0, t)$ (considered as a function of t) is not bounded.

1.3.5 The Cauchy problem for the nonhomogeneous wave equation

Consider the following Cauchy problem

$$u_{tt} - c^2 u_{xx} = F(x,t) \quad -\infty < x < \infty, t > 0, \tag{1.127}$$

$$u(x,0) = f(x), \quad u_t(x,0) = g(x) \quad -\infty < x < \infty. \tag{1.128}$$

This problem models, for example, the vibration of a very long string in the presence of an external force F. As in the homogeneous case f, g are given functions that represent the shape and the vertical velocity of the string at time $t = 0$.

As in every linear problem, the uniqueness for the homogeneous problem implies the uniqueness for the nonhomogeneous problem.

Proposition 1.3.2. *The Cauchy problem (1.127)-(1.128) admits at most one solution.*

Proof. Assume that u_1, u_2 are solutions of problem (1.127)-(1.128). We should prove that $u_1 = u_2$. The function $u = u_1 - u_2$ is a solution of the homogeneous problem

$$u_{tt} - c^2 u_{xx} = 0 \quad -\infty, < x < \infty, t > 0, \tag{1.129}$$

$$u(x,0) = 0, \quad u_t(x,0) = 0 \quad -\infty < x < \infty. \tag{1.130}$$

On the other hand, $v(x,t) = 0$ is also a solution of the same (homogeneous) problem. By Theorem 1.3.4, $u = v = 0$, hence, $u_1 = u_2$. □

Next, using an explicit formula, we prove, as in the homogeneous case, the existence of a solution of the Cauchy problem (1.127)-(1.128). For this purpose, recall Green's formula for a pair of functions P, Q in a planar domain Ω with a piecewise smooth boundary Γ :

$$\iint_{\Omega} [Q(x,t)_x - P(x,t)_t] \, \mathrm{d}x \, \mathrm{d}t = \oint_{\Gamma} [P(x,t)dx + Q(x,t)\mathrm{d}t].$$

Let $u(x,t)$ be a solution of problem (1.127)-(1.128). Integrate the two sides of the PDE (1.127) over a characteristic triangle $\triangle$ with a fixed upper vertex (x_0, t_0). The three edges of this triangle (base, right and left edges) will be denoted by B, R, L, respectively (see Figure 1.10). We have

$$-\iint_{\Delta} F(x,t)\mathrm{d}x \, \mathrm{d}t = \iint_{\Delta} \left(c^2 u_{xx} - u_{tt}\right) \mathrm{d}x \, \mathrm{d}t.$$

Using Green's formula with $Q = c^2 u_x$ and $P = u_t$, we obtain

$$\begin{aligned} -\iint_{\Delta} F(x,t)\mathrm{d}x \, \mathrm{d}t &= \oint_{\Gamma} \left(u_t \, \mathrm{d}x + c^2 u_x \, \mathrm{d}t\right) \\ &= \int_B + \int_R + \int_L \left(u_t \, \mathrm{d}x + c^2 u_x \, \mathrm{d}t\right). \end{aligned}$$

On the base B we have $\mathrm{d}t = 0$; therefore, using the initial conditions, we get

$$\int_B \left(u_t \, \mathrm{d}x + c^2 u_x \, \mathrm{d}t\right) = \int_{x_0 - ct_0}^{x_0 + ct_0} u_t(x, 0) \mathrm{d}x = \int_{x_0 - ct_0}^{x_0 + ct_0} g(x) \mathrm{d}x.$$

On the right edge $R, x + ct = x_0 + ct_0$, hence $\mathrm{d}x = -c \, \mathrm{d}t$. Consequently,

$$\int_R \left(u_t \, \mathrm{d}x + c^2 u_x \, \mathrm{d}t\right) = -c \int_R \left(u_t \, \mathrm{d}t + u_x \, \mathrm{d}x\right) = -c \int_R \mathrm{d}u$$
$$= -c \left[u\left(x_0, t_0\right) - u\left(x_0 + ct_0, 0\right)\right] = -c \left[u\left(x_0, t_0\right) - f\left(x_0 + ct_0\right)\right].$$

Similarly, on the left edge $L, x - ct = x_0 - ct_0$, implying $\mathrm{d}x = c \, \mathrm{d}t$, and

$$\int_L \left(u_t \, \mathrm{d}x + c^2 u_x \, \mathrm{d}t\right) = c \int_L \left(u_t \, \mathrm{d}t + u_x \, \mathrm{d}x\right) = c \int_L \mathrm{d}u$$
$$= c \left[u\left(x_0 - ct_0, 0\right) - u\left(x_0, t_0\right)\right] = c \left[f\left(x_0 - ct_0\right) - u\left(x_0, t_0\right)\right].$$

Therefore,

$$-\iint_\Delta F(x, t) \mathrm{d}x \, \mathrm{d}t = \int_{x_0 - ct_0}^{x_0 + ct_0} g(x) \mathrm{d}x$$
$$+ c \left[f\left(x_0 - ct_0\right) + f\left(x_0 + ct_0\right) - 2u\left(x_0, t_0\right)\right]$$

Solving for u gives

$$u\left(x_0,t_0\right)=\frac{f\left(x_0+ct_0\right)+f\left(x_0-ct_0\right)}{2}+\frac{1}{2c}\int_{x_0-ct_0}^{x_0+ct_0}g(x)\mathrm{d}x + \frac{1}{2c}\iint_{\Delta}F(x,t)\mathrm{d}x\ \mathrm{d}t.$$

We finally obtain an explicit formula for the solution at an arbitrary point (x,t) :

$$u(x,t)=\frac{f(x+ct)+f(x-ct)}{2}+\frac{1}{2c}\int_{x-ct}^{x+ct}g(s)\mathrm{d}s + \frac{1}{2c}\iint_{\Delta}F(\xi,\tau)\mathrm{d}\xi\mathrm{d}\tau. \tag{1.131}$$

This formula is also called d'Alembert's formula.

Remark 1.3.3. 1. Note that for $F=0$ the two d'Alembert's formulas coincide, and actually, we have obtained another proof of the d'Alembert formula (1.125).

2. The value of u at a point (x_0,t_0) is determined by the values of the given data on the whole characteristic triangle whose upper vertex is the point (x_0,t_0). This is the domain of dependence for the nonhomogeneous Cauchy problem.

It remains to prove that the function u in (1.131) is indeed a solution of the Cauchy problem. From the superposition principle it follows that u in (1.131) is the desired solution, if and only if the function

$$v(x,t) = \frac{1}{2c}\iint_{\Delta} F(\xi,\tau)\mathrm{d}\xi\mathrm{d}\tau = \frac{1}{2c}\int_0^t \int_{x-c(t-\tau)}^{x+c(t-\tau)} F(\xi,\tau)\mathrm{d}\xi\mathrm{d}\tau$$

is a solution of the Cauchy problem

$$u_{tt} - c^2 u_{xx} = F(x,t) \quad -\infty < x < \infty, t > 0, \tag{1.132}$$

$$u(x,0) = 0, \quad u_t(x,0) = 0 \quad -\infty < x < \infty. \tag{1.133}$$

We shall prove that v is a solution of the initial value problem (1.132)-(1.133) under the assumption that F and F_x are continuous. Clearly,

$$v(x,0) = 0.$$

In order to take derivatives, we shall use the formula

$$\frac{\partial}{\partial t}\int_{a(t)}^{b(t)} G(\xi,t)\mathrm{d}\xi = G(b(t),t)b'(t) - G(a(t),t)a'(t) + \int_{a(t)}^{b(t)} \frac{\partial}{\partial t}G(\xi,t)\mathrm{d}\xi.$$

Hence,

$$\begin{aligned} v_t(x,t) &= \frac{1}{2c}\int_x^x F(\xi,t)\mathrm{d}\xi + \\ &\qquad \frac{1}{2}\int_0^t [F(x+c(t-\tau),\tau) + F(x-c(t-\tau),\tau)]\mathrm{d}\tau \\ &= \frac{1}{2}\int_0^t [F(x+c(t-\tau),\tau) + F(x-c(t-\tau),\tau)]\mathrm{d}\tau. \end{aligned}$$

In particular,

$$v_t(x,0) = 0.$$

By taking the second derivative with respect to t, we have

$$v_{tt}(x,t) = F(x,t) + \frac{c}{2}\int_0^t [F_x(x+c(t-\tau),\tau) - F_x(x-c(t-\tau),\tau)]\,\mathrm{d}\tau$$

Similarly,

$$v_x(x,t) = \frac{1}{2c}\int_0^t [F(x+c(t-\tau),\tau) - F(x-c(t-\tau),\tau)]\mathrm{d}\tau,$$

$$v_{xx}(x,t) = \frac{1}{2c}\int_0^t [F_x(x+c(t-\tau),\tau) - F_x(x-c(t-\tau),\tau)]\,\mathrm{d}\tau.$$

Therefore, $v(x,t)$ is a solution of the nonhomogeneous wave equation (1.132), and the homogeneous initial conditions (1.133) are satisfied. Note that all the above differentiations are justified provided that $F, F_x \in C\left(\mathbb{R}^2\right)$.

Theorem 1.3.7. *Fix $T > 0$. The Cauchyproblem (1.127)-(1.128) in the domain $-\infty < x < \infty, 0 \le t \le T$ is well-posed for $F, F_x \in C\left(\mathbb{R}^2\right), f \in C^2(\mathbb{R}), g \in C^1(\mathbb{R})$.*

Proof. Recall that the uniqueness has already been proved, and the existence follows from d'Alembert's formula. It remains to prove stability, i.e. we need to show that small changes in the initial conditions and the

external force give rise to a small change in the solution. For $i = 1, 2$, let u_i be the solution of the Cauchy problem with the corresponding function F_i, and the initial conditions f_i, g_i. Now, if

$$|F_1(x,t) - F_2(x,t)| < \delta, \quad |f_1(x) - f_2(x)| < \delta, \quad |g_1(x) - g_2(x)| < \delta,$$

for all $x \in \mathbb{R}, 0 \le t \le T$, then for all $x \in \mathbb{R}$ and $0 \le t \le T$ we have

$$\begin{aligned} |u_1(x,t) - u_2(x,t)| \le \frac{|f_1(x+ct) - f_2(x+ct)|}{2} + \frac{|f_1(x-ct) - f_2(x-ct)|}{2} \\ + \frac{1}{2c}\int_{x-ct}^{x+ct} |g_1(s) - g_2(s)|\,\mathrm{d}s + \\ \frac{1}{2c}\iint_{\Delta} |F_1(\xi,\tau) - F_2(\xi,\tau)|\,\mathrm{d}\xi\mathrm{d}\tau \end{aligned}$$

$$< \frac{1}{2}(\delta + \delta) + \frac{1}{2c} 2ct\delta + \frac{1}{2c} ct^2\delta \le \left(1 + T + T^2/2\right)\delta.$$

Therefore, for a given $\varepsilon > 0$, we choose $\delta < \varepsilon/\left(1 + T + T^2/2\right)$. Thus, for all $x \in \mathbb{R}$ and $0 \le t \le T$, we have

$$|u_1(x,t) - u_2(x,t)| < \varepsilon.$$

Note that δ does not depend on the wave speed c. □

Corollary 1.3.4. *Suppose that f, g are even functions, and for every $t \ge 0$ the function $F(\cdot, t)$ is even too. Then for every $t \ge 0$ the solution $u(\cdot, t)$ of the Cauchy problem (1.127)-(1.128) is also even. Similarly, the*

solution is an odd function or a periodic function with a period L (as a function of x) if the data are odd functions or periodic functions with a period L.

Proof. We prove the first part of the corollary. The other parts can be shown similarly. Let u be the solution of the problem and define the function $v(x,t) = u(-x,t)$. Clearly,

$$v_x(x,t) = -u_x(-x,t), \quad v_t(x,t) = u_t(-x,t)$$

and

$$v_{xx}(x,t) = u_{xx}(-x,t), \quad v_{tt}(x,t) = u_{tt}(-x,t).$$

Therefore,

$$\begin{aligned} v_{tt}(x,t) - c^2 v_{xx}(x,t) &= u_{tt}(-x,t) - c^2 u_{xx}(-x,t) \\ &= F(-x,t) = F(x,t) \quad -\infty < x < \infty, t > 0. \end{aligned}$$

Thus, v is a solution of the nonhomogeneous wave equation (1.127). Furthermore,

$v(x,0) = u(-x,0) = f(-x) = f(x)$, $v_t(x,0) = u_t(-x,0) = g(-x) = g(x)$.

It means that v is also a solution of the initial value problem (1.127)-(1.128). Since the solution of this problem is unique, we have $v(x,t) = u(x,t)$, which implies $u(-x,t) = u(x,t)$. □

Example 1.3.8. Solve the following Cauchy problem

$$\begin{aligned} &u_{tt} - 9u_{xx} = \mathrm{e}^{x} - \mathrm{e}^{-x} && -\infty < x < \infty, t > 0, \\ &u(x,0) = x && -\infty < x < \infty, \\ &u_t(x,0) = \sin x && -\infty < x < \infty. \end{aligned}$$

Using the d'Alembert formula, we have

$$\begin{aligned} u(x,t) =& \frac{1}{2}[f(x+ct) + f(x-ct)] + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)\mathrm{d}s \\ &+ \frac{1}{2c}\int_{\tau=0}^{\tau=t}\int_{\xi=x-c(t-\tau)}^{\xi=x+c(t-\tau)} F(\xi,\tau)\mathrm{d}\xi\mathrm{d}\tau. \end{aligned}$$

Hence,

$$\begin{aligned} u(x,t) =& \frac{1}{2}[x+3t+x-3t] + \frac{1}{6}\int_{x-3t}^{x+3t} \sin s \ \mathrm{d}s \\ &+ \frac{1}{6}\int_{\tau=0}^{\tau=t}\int_{\xi=x-3(t-\tau)}^{\xi=x+3(t-\tau)} \left(\mathrm{e}^{\xi} - \mathrm{e}^{-\xi}\right)\mathrm{d}\xi\mathrm{d}\tau \\ =& x + \frac{1}{3}\sin x \sin 3t - \frac{2}{9}\sinh x + \frac{2}{9}\sinh x \cosh 3t. \end{aligned}$$

As expected, for all $t \geq 0$, the solution u is an odd function of x.

Remark 1.3.5. In many cases it is possible to reduce a nonhomogeneous problem to a homogeneous problem if we can find a particular solution v of the given nonhomogeneous equation. This will eliminate the need to perform the double integration which appears in the d'Alembert formula (1.131). The technique is particularly useful when F has a simple form, for example, when $F = F(x)$ or $F = F(t)$. Suppose that such a particular solution v is found, and consider the function $w = u - v$. By

the superposition principle, w should solve the following homogeneous Cauchy problem:

$$\begin{aligned} &w_{tt} - w_{xx} = 0 && -\infty < x < \infty, t > 0, \\ &w(x,0) = f(x) - v(x,0) && -\infty < x < \infty, \\ &w_t(x,0) = g(x) - v_t(x,0) && -\infty < x < \infty. \end{aligned}$$

Hence, w can be found using the d'Alembert formula for the homogeneous equation. Then $u = v + w$ is the solution of our original problem.

We illustrate this idea through the following example.

Example 1.3.9. Solve the problem

$$\begin{aligned} &u_{tt} - u_{xx} = t^7 && -\infty < x < \infty, t > 0, \\ &u(x,0) = 2x + \sin x && -\infty < x < \infty, \\ &u_t(x,0) = 0 && -\infty < x < \infty. \end{aligned}$$

Because of the special form of the nonhomogeneous equation, we look for a particular solution of the form $v = v(t)$. Indeed it can be easily verified that $v(x,t) = \frac{1}{72}t^9$ is such a solution. Consequently, we need to solve the homogeneous problem

$$\begin{aligned} &w_{tt} - w_{xx} = 0 && -\infty < x < \infty, t > 0, \\ &w(x,0) = f(x) - v(x,0) = 2x + \sin x && -\infty < x < \infty, \\ &w_t(x,0) = g(x) - v_t(x,0) = 0 && -\infty < x < \infty. \end{aligned}$$

Using d'Alembert's formula for the homogeneous equation, we have

$$w(x,t) = 2x + \frac{1}{2}\sin(x+t) + \frac{1}{2}\sin(x-t),$$

and the solution of the original problem is given by

$$u(x,t) = 2x + \sin x \cos(t) + \frac{t^9}{72}.$$

Exercises

1. Complete the proof of Corollary 1.3.4.

2. Solve the problem

$$\begin{aligned} u_{tt} - u_{xx} &= 0 \quad && 0 < x < \infty, t > 0, \\ u(0,t) &= t^2 && t > 0, \\ u(x,0) &= x^2 && 0 \le x < \infty, \\ u_t(x,0) &= 6x && 0 \le x < \infty, \end{aligned}$$

   and evaluate $u(4,1)$ and $u(1,4)$.

3. Consider the problem

$$u_{tt} - 4u_{xx} = 0 \quad -\infty < x < \infty, t > 0,$$

$$u(x,0) = \begin{cases} 1 - x^2 & |x| \le 1, \\ 0 & \text{otherwise} \end{cases}$$

$$u_t(x,0) = \begin{cases} 4 & 1 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Using the graphical method, find $u(x,1)$.

(b) Find $\lim_{t\to\infty} u(5,t)$.

(c) Find the set of all points where the solution is singular (non-classical).

(d) Find the set of all points where the solution is not continuous.

4. (a) Solve the following initial boundary value problem for a vibrating semi-infinite string which is fixed at $x = 0$:

$$\begin{aligned} u_{tt} - u_{xx} &= 0 \qquad && 0 < x < \infty, t > 0, \\ u(0,t) &= 0 && t > 0, \\ u(x,0) &= f(x) && 0 \le x < \infty, \\ u_t(x,0) &= g(x) && 0 \le x < \infty, \end{aligned}$$

where $f \in C^2([0,\infty))$ and $g \in C^1([0,\infty))$ satisfy the compatibility conditions $f(0) = f''(0) = g(0) = 0$.

Hint: Extend the functions f and g as odd functions $\tilde{f}$ and $\tilde{g}$ over the real line. Solve the Cauchy problem with initial data $\tilde{f}$ and $\tilde{g}$, and show that the restriction of this solution to the half-plane $x \ge 0$ is a solution of the problem. Recall that the solution of the Cauchy problem with odd data is odd. In particular, the solution with odd data is zero for $x = 0$ and all $t \ge 0$.

(b) Solve the problem with $f(x) = x^3 + x^6$, and $g(x) = \sin^2 x$, and evaluate $u(1,i)$ for $i = 1, 2, 3$. Is the solution classical?

5. Consider the problem

$$u_{tt} - u_{xx} = 0 \quad -\infty < x < \infty, t > 0,$$

$$u(x,0) = \begin{cases} 8x - 2x^2 & 0 \le x \le 4, \\ 0 & \text{otherwise}, \end{cases}$$

$$u_t(x,0) = \begin{cases} 16 & 0 \le x \le 4, \\ 0 & \text{otherwise}. \end{cases}$$

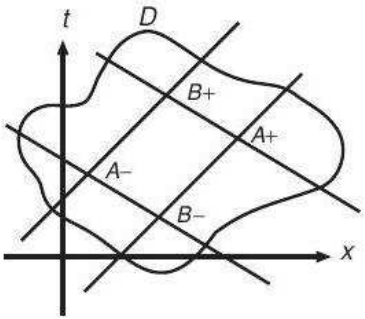

Figure 1.14: A drawing for the parallelogram identity.

(a) Find a formula for the forward and backward waves.

(b) Using the graphical method, draw the graph of $u(x,i)$ for $i = 4, 8, 12$.

(c) Find $u(\pm 5, 2), u(\pm 3, 4)$.

(d) Find $\lim_{t\to\infty} u(5,t)$.

6. (a) Solve the following initial boundary value problem for a vibrating semi-infinite string with a free boundary condition:

$$\begin{aligned} u_{tt} - u_{xx} &= 0 \qquad && 0 < x < \infty, t > 0, \\ u_x(0,t) &= 0 && t > 0, \\ u(x,0) &= f(x) && 0 \le x < \infty, \\ u_t(x,0) &= g(x) && 0 \le x < \infty, \end{aligned}$$

where $f \in C^2([0,\infty))$ and $g \in C^1([0,\infty))$ satisfy the compatibility conditions $f'_+(0) = g'_+(0) = 0$.

Hint: Extend the functions f and g as even functions $\tilde{f}$ and $\tilde{g}$ on the line. Solve the Cauchy problem with initial data $\tilde{f}$ and $\tilde{g}$, and show that the restriction of this solution to the half-plane $x \ge 0$ is a solution of the problem.

(b) Solve the problem with $f(x) = x^3 + x^6, g(x) = \sin^3 x$, and evaluate $u(1,i)$ for $i = 1,2,3$. Is the solution classical?

7. (a) Let $u(x,t)$ be a solution of the wave equation $u_{tt} - u_{xx} = 0$ in a domain $D \subset \mathbb{R}^2$. Let a, b be real numbers such that the parallelogram with vertices $A_\pm = (x_0 \pm a, t_0 \pm b)$, $B_\pm = (x_0 \pm b, t_0 \pm a)$ is contained in D (see Figure 1.14). Prove the parallelogram identity:

$$u(x_0 - a, t_0 - b) + u(x_0 + a, t_0 + b) = u(x_0 - b, t_0 - a) + u(x_0 + b, t_0 + a).$$

(b) Derive the corresponding identity when the wave speed $c \neq 1$.

(c) Using the parallelogram identity, solve the following initial boundary value problem for a vibrating semi-infinite string with a nonhomogeneous boundary condition:

$$\begin{aligned} u_{tt} - u_{xx} &= 0 \qquad 0 < x < \infty, t > 0, \\ u(0,t) &= h(t) \quad t > 0, \\ u(x,0) &= f(x) \quad 0 \leq x < \infty, \\ u_t(x,0) &= g(x) \quad 0 \leq x < \infty, \end{aligned}$$

where $f, g, h \in C^2([0,\infty))$. Hint Distinguish between the cases $x - t > 0$ and $x - t \leq 0$.

(d) From the explicit formula that was obtained in part (c), derive the corresponding compatibility conditions, and prove that the problem is well-posed.

(e) Derive an explicit formula for the solution and deduce the corresponding compatibility conditions for the case $c \neq 1$.

8. Solve the following initial boundary value problem using the parallelogram identity

$$\begin{aligned} u_{tt} - u_{xx} &= 0 \qquad 0 < x < \infty, 0 < t < 2x, \\ u(x,0) &= f(x) \quad 0 \leq x < \infty, \\ u_t(x,0) &= g(x) \quad 0 \leq x < \infty, \\ u(x,2x) &= h(x) \quad x \geq 0, \end{aligned}$$

where $f, g, h \in C^2([0,\infty))$.

9. Solve the problem

$$\begin{aligned} u_{tt} - u_{xx} &= 1 \qquad -\infty < x < \infty, t > 0 \\ u(x,0) &= x^2 \quad -\infty < x < \infty \\ u_t(x,0) &= 1 \quad -\infty < x < \infty \end{aligned}$$

10. (a) Solve the Darboux problem:

$$u_{tt} - u_{xx} = 0 \qquad t > \max\{-x, x\}, t \geq 0,$$
$$u(x,t) = \begin{cases} \phi(t) & x = t, t \geq 0, \\ \psi(t) & x = -t, t \geq 0, \end{cases}$$

where $\phi, \psi \in C^2([0,\infty)$ satisfies $\phi(0) = \psi(0)$.

(b) Prove that the problem is well posed.

11. A pressure wave generated as a result of an explosion satisfies the equation

$$P_{tt} - 16P_{xx} = 0$$

in the domain $\{(x,t) \mid -\infty < x < \infty, t > 0\}$, where $P(x,t)$ is the pressure at the point x and time t. The initial conditions at the explosion time $t = 0$ are

$$P(x,0) = \begin{cases} 10 & |x| \leq 1, \\ 0 & |x| > 1, \end{cases}$$
$$P_t(x,0) = \begin{cases} 1 & |x| \leq 1, \\ 0 & |x| > 1. \end{cases}$$

A building is located at the point $x_0 = 10$. The engineer who designed the building determined that it will sustain a pressure up to $P = 6$. Find the time t_0 when the pressure at the building is maximal. Will the building collapse?

12. (a) Solve the problem

$$\begin{aligned} u_{tt} - u_{xx} &= 0 && 0 < x < \infty, 0 < t, \\ u(0,t) &= \frac{t}{1+t} && 0 \le t, \\ u(x,0) = u_t(x,0) &= 0 && 0 \le x < \infty. \end{aligned}$$

(b) Show that the limit

$$\lim_{x\to\infty} u(cx, x) := \phi(c)$$

exists for all $c > 0$. What is the limit?

13. Consider the Cauchy problem

$$\begin{aligned} u_{tt} - 4u_{xx} &= F(x,t) \quad -\infty < x < \infty, t > 0, \\ u(x,0) = f(x), \quad u_t(x,0) &= g(x) \quad -\infty < x < \infty, \end{aligned}$$

where

$$f(x) = \begin{cases} x & 0 < x < 1, \\ 1 & 1 < x < 2, \\ 3 - x & 2 < x < 3, \\ 0 & x > 3, x < 0, \end{cases}$$

$$g(x) = \begin{cases} 1 - x^2 & |x| < 1, \\ 0 & |x| > 1, \end{cases}$$

and $F(x,t) = -4\mathrm{e}^x$ on $t > 0, -\infty < x < \infty$.

(a) Is the d'Alembert solution of the problem a classical solution? If your answer is negative, find all the points where the solution is singular.

(b) Evaluate the solution at $(1, 1)$.

14. Solve the problem

$$\begin{aligned} u_{tt} - 4u_{xx} &= \mathrm{e}^x + \sin t \quad && -\infty < x < \infty, t > 0, \\ u(x, 0) &= 0 && -\infty < x < \infty, \\ u_t(x, 0) &= \tfrac{1}{1+x^2} && -\infty < x < \infty. \end{aligned}$$

15. Find the general solution of the problem

$$u_{ttx} - u_{xxx} = 0, \quad u_x(x, 0) = 0, \quad u_{xt}(x, 0) = \sin x,$$

in the domain $\{(x, t) \mid -\infty < x < \infty, \quad t > 0\}$.

16. Solve the problem

$$\begin{aligned} u_{tt} - u_{xx} &= xt \quad && -\infty < x < \infty, t > 0 \\ u(x, 0) &= 0 && -\infty < x < \infty \\ u_t(x, 0) &= \mathrm{e}^x && -\infty < x < \infty \end{aligned}$$

17. (a) Without using the d'Alembert formula find a solution $u(x, t)$ of the problem

$$\begin{aligned} u_{tt} - u_{xx} &= \cos(x + t) \quad && -\infty < x < \infty, t > 0 \\ u(x, 0) &= x, \quad u_t(x, 0) = \sin x && -\infty < x < \infty \end{aligned}$$

(b) Without using the d'Alembert formula find $v(x, t)$ that is a solution of the problem

$$\begin{aligned} v_{tt} - v_{xx} &= \cos(x + t) \quad && -\infty < x < \infty, t > 0, \\ v(x, 0) &= 0, \quad v_t(x, 0) = 0 && -\infty < x < \infty. \end{aligned}$$

(c) Find the PDE and initial conditions that are satisfied by the function $w := u - v$.

(d) Which of the functions u, v, w (as a function of x) is even? Odd? Periodic?

(e) Evaluate $v(2\pi, \pi), w(0, \pi)$.

18. Solve the problem

$$\begin{aligned} u_{tt} - 4u_{xx} = 6t \quad & -\infty < x < \infty, t > 0, \\ u(x, 0) = x \quad & -\infty < x < \infty, \\ u_t(x, 0) = 0 \quad & -\infty < x < \infty, \end{aligned}$$

without using the d'Alembert formula.

19. Let $u(x, t)$ be a solution of the equation $u_{tt} - u_{xx} = 0$ in the whole plane. Suppose that $u_x(x, t)$ is constant on the line $x = 1 + t$. Assume also that $u(x, 0) = 1$ and $u(1, 1) = 3$. Find such a solution $u(x, t)$. Is this solution uniquely determined?

Module 2

The Method of Separation of Variables and Elliptic Equations

2.1 The method of separation of variables

2.1.1 Introduction

We examined in Module 1 Fourier's work on heat conduction. In addition to developing a general theory for heat flow, Fourier discovered a method for solving the initial boundary value problem he derived. His solution led him to propose the bold idea that any real valued function defined on a closed interval can be represented as a series of trigonomet-

ric functions. This is known today as the Fourier expansion. D'Alembert and the Swiss mathematician Daniel Bernoulli (1700-1782) had actually proposed a similar idea before Fourier. They claimed that the vibrations of a finite string can be formally represented as an infinite series involving sinusoidal functions. They failed, however, to see the generality of their observation.

Fourier's method for solving the heat equation provides a convenient method that can be applied to many other important linear problems. The method also enables us to deduce several properties of the solutions, such as asymptotic behavior, smoothness, and well-posedness. Historically, Fourier's idea was a breakthrough which paved the way for new developments in science and technology. For example, Fourier analysis found many applications in pure mathematics (number theory, approximation theory, etc.). Several fundamental theories in physics (quantum mechanics in particular) are heavily based on Fourier's idea, and the entire theory of signal processing is based on Fourier's method and its generalizations.

2.1.2 Heat equation: homogeneous boundary conditions

Consider the following heat conduction problem in a finite interval:

$$u_t - ku_{xx} = 0 \quad 0 < x < L, t > 0, \tag{2.1}$$

$$u(0,t) = u(L,t) = 0 \quad t \geq 0, \tag{2.2}$$

$$u(x,0) = f(x) \quad 0 \leq x \leq L, \tag{2.3}$$

where f is a given initial condition, and k is a positive constant. In order to make (2.2) consistent with (2.3), we assume the compatibility condition

$$f(0) = f(L) = 0$$

The equation and the domain are drawn schematically in Figure 2.1.

The problem defined above corresponds to the evolution of the temperature $u(x,t)$ in a homogeneous one-dimensional heat conducting rod of length L (i.e. the rod is narrow and is laterally insulated) whose initial temperature (at time $t = 0$) is known and is such that its two ends are immersed in a zero temperature bath.

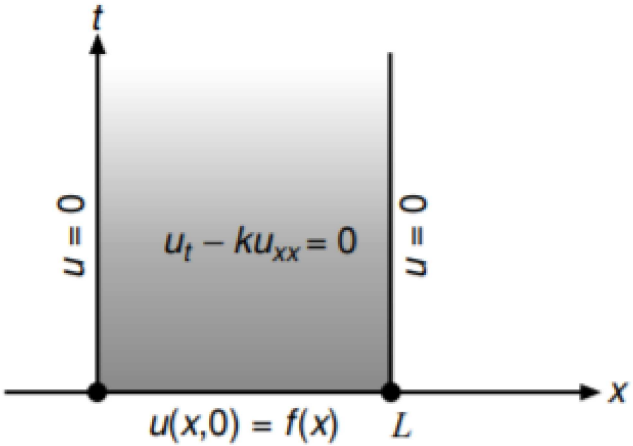

Figure 2.1: The initial boundary value problem for the heat equation together with the domain.

We assume that there is no internal source that heats (or cools) the system. Note that the problem (2.1) – (2.3) is an initial boundary value

problem that is linear and homogeneous. Recall also that the boundary condition (2.2) is called the *Dirichlet condition*. At the end of the present section, we shall also discuss other boundary conditions.

$$U(x,t) = X(x)T(t), \tag{2.4}$$

where X and T are functions of the variables x and t, respectively. At this step we do not take into account the initial condition (2.3). Obviously, we are not interested in the zero solution $u(x,t) = 0$. Therefore, we seek functions X and T that do not vanish identically.

Differentiate the separated solution (2.4) once with respect to t and twice with respect to x and substitute these derivatives into the PDE. We then obtain

$$XT_t = kX_{xx}T.$$

Now, we carry out a simple but decisive step - the *separation of variables step*. We move to one side of the PDE all the functions that depend only on x and to the other side the functions that depend only on t. We thus write

$$\frac{T_t}{kT} = \frac{X_{xx}}{X}. \tag{2.5}$$

Since x and t are independent variables, differentiating (2.5) with respect to t implies that there exists a constant denoted by λ (which is called the separation constant) such that

$$\frac{T_t}{kT} = \frac{X_{xx}}{X} = -\lambda \tag{2.6}$$

Equation (2.6) leads to the following system of ODEs:

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = -\lambda X \quad 0 < x < L, \tag{2.7}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\lambda kT \quad t > 0, \tag{2.8}$$

which are coupled only by the separation constant λ. The function u satisfies the boundary conditions (2.2) if and only if

$$u(0,t) = X(0)T(t) = 0, \quad u(L,t) = X(L)T(t) = 0.$$

Since u is not the trivial solution $u = 0$, it follows that

$$X(0) = X(L) = 0.$$

Therefore, the function X should be a solution of the boundary value problem

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2} + \lambda X = 0 \quad 0 < x < L, \tag{2.9}$$

$$X(0) = X(L) = 0. \tag{2.10}$$

Consider the system (2.9)-(2.10). A nontrivial solution of this system is called an eigenfunction of the problem with an eigenvalue λ. The problem (2.9)-(2.10) is called an eigenvalue problem. The boundary condition (2.10) is called (as in the PDE case) the Dirichlet boundary condition.

Note that the problem (2.9)-(2.10) is not an initial boundary problem for an ODE (for which it is known that there exists a unique solution). Rather, it is a boundary value problem for an ODE. It is not clear a priori that there exists a solution for any value of λ. On the other hand, if we can write the general solution of the ODE for every λ, then we

need only to check for which λ there exists a solution that also satisfies the boundary conditions.

Fortunately, (2.9) is quite elementary. It is a second-order linear ODE with constant coefficients, and its general solution (which depends on λ) has the following form:

1. if $\lambda < 0$, then $X(x) = \alpha e^{\sqrt{-\lambda}x} + \beta e^{-\sqrt{-\lambda}x}$,
2. if $\lambda = 0$, then $X(x) = \alpha + \beta x$,
3. if $\lambda > 0$, then $X(x) = \alpha \cos(\sqrt{\lambda}x) + \beta \sin(\sqrt{\lambda}x)$,

where α, β are arbitrary real numbers.

We implicitly assume that λ is real, and we do not consider the complex case (although this case can, in fact, be treated similarly). We can show that the system (2.9)-(2.10) does not admit a solution with a nonreal λ. In other words, all the eigenvalues of the problem are real numbers.

Negative eigenvalue ($\lambda < 0$) The general solution can be written in a more convenient form: instead of choosing the two exponential functions as the fundamental system of solutions, we use the basis $\{\sinh(\sqrt{-\lambda}x), \cosh(\sqrt{-\lambda}x)\}$. In this basis, the general solution for $\lambda < 0$ has the form

$$X(x) = \tilde{\alpha} \cosh(\sqrt{-\lambda}x) + \tilde{\beta} \sinh(\sqrt{-\lambda}x). \tag{2.11}$$

The function $\sinh s$ has a unique root at $s = 0$, while $\cosh s$ is a strictly positive function. Since $X(x)$ should satisfy $X(0) = 0$, it follows $\tilde{\alpha} = 0$. The second boundary condition $X(L) = 0$ implies that $\tilde{\beta} = 0$.

Hence, $X(x) \equiv 0$ is the trivial solution. In other words, the system (2.9)-(2.10) does not admit a negative eigenvalue.

Zero eigenvalue ($\lambda = 0$) We claim that $\lambda = 0$ is also not an eigenvalue. Indeed, in this case the general solution is a linear function $X(x) = \alpha + \beta x$ that (in the nontrivial case $X \neq 0$) vanishes at most at one point; thus it cannot satisfy the boundary conditions (2.10).

Positive eigenvalue ($\lambda > 0$) The general solution for $\lambda > 0$ is

$$X(x) = \alpha \cos(\sqrt{\lambda}x) + \beta \sin(\sqrt{\lambda}x). \tag{2.12}$$

Substituting this solution into the boundary condition $X(0) = 0$, we obtain $\alpha = 0$. The boundary condition $X(L) = 0$ implies $\sin(\sqrt{\lambda}L) = 0$. Therefore, $\sqrt{\lambda}L = n\pi$, where n a positive integer. We do not have to consider the case $n < 0$, since it corresponds to the same set of eigenvalues and eigenfunctions. Hence, λ is an eigenvalue if and only if

$$\lambda = \left(\frac{n\pi}{L}\right)^2 \quad n = 1, 2, 3, \ldots$$

The corresponding eigenfunctions are

$$X(x) = \sin \frac{n\pi x}{L},$$

and they are uniquely defined up to a multiplicative constant.

In conclusion, the set of all solutions of problem (2.9)-(2.10) is an infinite sequence of eigenfunctions, each associated with a positive eigenvalue. It is convenient to use the notation

$$X_n(x) = \sin \frac{n\pi x}{L}, \quad \lambda_n = \left(\frac{n\pi}{L}\right)^2 \quad n = 1, 2, 3, \ldots$$

Recall from linear algebra that an eigenvalue has multiplicity m if the space consisting of its eigenvectors is m-dimensional. An eigenvalue with multiplicity 1 is called simple. Using the same terminology, we see that the eigenvalues λ_n for the eigenvalue problem (2.9)-(2.10) are all simple. Let us deal now with the ODE (2.8). The general solution has the form

$$T(t) = B\mathrm{e}^{-k\lambda t}.$$

Substituting λ_n, we obtain

$$T_n(t) = B_n\mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \quad n = 1, 2, 3, \ldots \tag{2.13}$$

From the physical point of view it is clear that the solution of (2.8) must decay in time, hence, we must have $\lambda > 0$. Therefore, we could have guessed a priori that the problem (2.9)-(2.10) would admit only positive eigenvalues.

We have thus obtained the following sequence of separated solutions

$$u_n(x,t) = X_n(x)T_n(t) = B_n \sin\frac{n\pi x}{L}\mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \quad n = 1, 2, 3, \ldots \tag{2.14}$$

The superposition principle implies that any linear combination

$$u(x,t) = \sum_{n=1}^{N} B_n \sin\frac{n\pi x}{L}\mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \tag{2.15}$$

of separated solutions is also a solution of the heat equation that satisfies the Dirichlet boundary conditions.

Consider now the initial condition. Suppose it has the form

$$f(x) = \sum_{n=1}^{N} B_n \sin \frac{n\pi x}{L},$$

i.e. it is a linear combination of the eigenfunctions. Then a solution of the heat problem (2.1)-(2.3) is given by

$$u(x,t) = \sum_{n=1}^{N} B_n \sin \frac{n\pi x}{L} \mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t}.$$

Hence, we are able to solve the problem for a certain family of initial conditions.

It is natural to ask at this point how to solve for more general initial conditions? The brilliant (although not fully justified at that time) idea of Fourier was that it is possible to represent an arbitrary function f that satisfies the boundary conditions (2.2) as a unique infinite "linear combination" of the eigenfunctions $\sin(n\pi x/L)$. In other words, it is possible to find constants B_n such that

$$f(x) = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L}. \tag{2.16}$$

Such a series is called a (generalized) Fourier series (or expansion) of the function f with respect to the eigenfunctions of the problem, and $B_n, n = 1, 2 \ldots$ are called the (generalized) Fourier coefficients of the series. The last ingredient that is needed for solving the problem is called the generalized superposition principle. We generalize the superposition principle and apply it also to an infinite series of separated solutions. We call such a series a generalized solution of the PDE if the series

is uniformly converging in every subrectangle that is contained in the domain where the solution is defined. This definition is similar to the definition of generalized solutions of the wave equation that was given in Module 1.

In our case the generalized superposition principle implies that the formal expression

$$u(x,t) = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L} \mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \tag{2.17}$$

is a natural candidate for a generalized solution of problem (2.1)-(2.3). By a 'formal solution' we mean that if we ignore questions concerning convergence, continuity, and smoothness, and carry out term-by-term differentiations and substitutions, then we see that all the required conditions of the problem (2.1)-(2.3) are satisfied.

Before proving that under certain conditions (2.17) is indeed a solution, we need to explain how to represent an 'arbitrary' function f as a Fourier series. In other words, we need a method of finding the Fourier coefficients of a given function f.

Surprisingly, this question can easily be answered under the assumption that the Fourier series of f converges uniformly. Fix $m \in \mathbb{N}$, multiply the Fourier expansion (2.16) by the eigenfunction $\sin(m\pi x/L)$, and then integrate the equation term-byterm over $[0, L]$. We get

$$\int_0^L \sin \frac{m\pi x}{L} f(x) \mathrm{d}x = \sum_{n=1}^{\infty} B_n \int_0^L \sin \frac{m\pi x}{L} \sin \frac{n\pi x}{L} \, \mathrm{d}x. \tag{2.18}$$

It is easily checked that

$$\int_0^L \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}\,\mathrm{d}x = \begin{cases} 0 & m \neq n \\ L/2 & m = n \end{cases} \tag{2.19}$$

Therefore, the Fourier coefficients are given by

$$\begin{aligned} B_m &= \frac{\int_0^L \sin(m\pi x/L) f(x)\mathrm{d}x}{\int_0^L \sin^2(m\pi x/L)\mathrm{d}x} \\ &= \tfrac{2}{L}\int_0^L \sin\tfrac{m\pi x}{L} f(x)\mathrm{d}x, \end{aligned} \tag{2.20}$$

Where $m = 1, 2, 3, \ldots$

In particular, it follows that the Fourier coefficients and the Fourier expansion of f are uniquely determined. Therefore, (2.17) together with (2.20) provides an explicit formula for a (formal) solution of the heat problem. Notice that we have developed a powerful tool! For a given initial condition f, one only has to compute the corresponding Fourier coefficients in order to obtain an explicit solution.

Example 2.1 Consider the problem:

$$u_t - u_{xx} = 0 \quad 0 < x < \pi, t > 0, \tag{2.21}$$

$$u(0,t) = u(\pi,t) = 0 \quad t \geq 0, \tag{2.22}$$

$$u(x,0) = f(x) = \begin{cases} x & 0 \leq x \leq \pi/2, \\ \pi - x & \pi/2 \leq x \leq \pi. \end{cases} \tag{2.23}$$

The formal solution is

$$u(x,t) = \sum_{m=1}^{\infty} B_m \sin mx \mathrm{e}^{-m^2 t}, \tag{2.24}$$

where

$$\begin{aligned} B_m &= \tfrac{2}{\pi}\int_0^{\pi} f(x)\sin mx\ \mathrm{d}x \\ &= \tfrac{2}{\pi}\int_0^{\pi/2} x\sin mx\ \mathrm{d}x + \tfrac{2}{\pi}\int_{\pi/2}^{\pi}(\pi-x)\sin mx\ \mathrm{d}x \\ &= \tfrac{2}{\pi}\left[\tfrac{-x\cos mx}{m} + \tfrac{\sin mx}{m^2}\right]_0^{\pi/2} \\ &\quad + \tfrac{2}{\pi}\left[\tfrac{-(\pi-x)\cos mx}{m} - \tfrac{\sin mx}{m^2}\right]_{\pi/2}^{\pi} \\ &= \tfrac{4}{\pi m^2}\sin\tfrac{m\pi}{2} \end{aligned} \tag{2.25}$$

But

$$\sin\frac{m\pi}{2} = \begin{cases} 0 & m = 2n, \\ (-1)^{n+1} & m = 2n-1, \end{cases} \tag{2.26}$$

where $n = 1, 2, \ldots$. Therefore, the formal solution is

$$\begin{aligned} u(x,t) &= \textstyle\sum_{n=1}^{\infty} u_n(x,t) \\ &= \tfrac{4}{\pi}\textstyle\sum_{n=1}^{\infty} \tfrac{(-1)^{n+1}}{(2n-1)^2}\sin[(2n-1)x]\mathrm{e}^{-(2n-1)^2 t} \end{aligned} \tag{2.27}$$

We claim that under the assumption that the Fourier expansion converges to f, the series (2.27) is indeed a classical solution. To verify this statement we assume

$$f(x) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{(2n-1)^2}\sin[(2n-1)x]. \tag{2.28}$$

The functions obtained by summing only finitely many terms in the Fourier series are depicted in Figure 2.2.

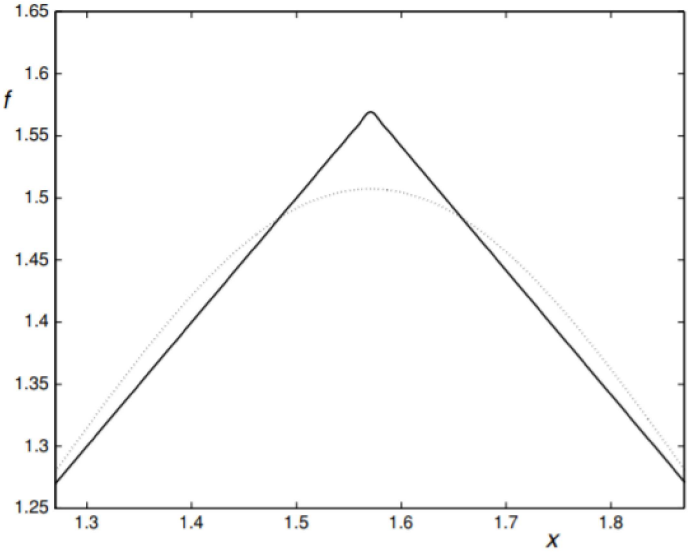

Figure 2.2: The function obtained by summing 100 terms (solid line) and 7 terms (dotted line) for the Fourier expansion of $f(x)$(5.27). We concentrate the region near the point $x = \pi/2$, where $f(x)$ is not differentiable. If we take just a few terms in the expansion, the actual singularity is smoothed out.

Since

$$|u_n(x,t)| = \frac{4}{\pi}\left|\frac{(-1)^{n+1}}{(2n-1)^2}\sin[(2n-1)x]\mathrm{e}^{-(2n-1)^2 t}\right| \leq \frac{4}{\pi(2n-1)^2},$$

it follows by the Weierstrass M-test that the series (2.27) converges uniformly to a continuous function in the region

$$\{(x,t) \mid 0 \leq x \leq \pi, t \geq 0\}.$$

Substituting u into the initial and boundary conditions, and using the assumption that the Fourier expansion of f converges to f, we obtain that these conditions are indeed satisfied.

It remains to show that the series (2.27) is differentiable with respect to t, twice differentiable with respect to x, and satisfies the heat equation in the domain

$$D := \{(x,t) \mid 0 < x < \pi, t > 0\}.$$

Fix $\varepsilon > 0$. We first show that the series (2.27) is differentiable with respect to t, twice differentiable with respect to x, and satisfies the heat equation in the subdomain

$$D_\varepsilon := \{(x,t) \mid 0 < x < \pi, t > \varepsilon\}$$

For instance, we show that (2.27) can be differentiated with respect to t for $t > \varepsilon$. Indeed, by differentiating $u_n(x,t)$ with respect to t, we obtain that

$$|(u_n(x,t))_t| = \left| \frac{4(2n-1)^2}{\pi(2n-1)^2} \sin[(2n-1)x] \mathrm{e}^{-(2n-1)^2 t} \right| \leq \frac{4}{\pi} \mathrm{e}^{-(2n-1)^2 \varepsilon}.$$

Since the series $(4/\pi) \sum \mathrm{e}^{-(2n-1)^2 \varepsilon}$ converges, it follows by the Weierstrass Mtest that for every $\varepsilon > 0$ the series $\sum (u_n(x,t))_t$ converges to u_t uniformly in D_ε. Similarly, it can be shown that u has a continuous second-order derivative with respect to x that is obtained by two term-by-term differentiations. Hence,

$$u_t - u_{xx} = \sum_{n=1}^{\infty} (u_n)_t - \sum_{n=1}^{\infty} (u_n)_{xx} = \sum_{n=1}^{\infty} \{(u_n)_t - (u_n)_{xx}\} = 0$$

where in the last step we used the property that each separated solution

$u_n(x,t)$ is a solution of the heat equation. Thus, u is a solution of the PDE in D_ε. Since ε is an arbitrary positive number, it follows that u is a solution of the heat equation in the domain D.

Because the general term u_n decays exponentially in D_ε, it is possible to differentiate (2.27) term-by-term to any order with respect to x and t. The corresponding series converges uniformly in D_ε to the appropriate derivative. Note that k differentiations with respect to x and ℓ differentiations with respect to t contribute to the general term of the series a factor of order $O\left(n^{k+2\ell}\right)$, but because of the exponential term, the corresponding series is converging.

The important conclusion is that even for nonsmooth initial condition f, the solution has infinitely many derivatives with respect to x and t and it is smooth in the strip D. The nonsmoothness of the initial data disappears immediately (see Figure 2.3). This smoothing effect is known to hold also in more general parabolic problems, in contrast with the hyperbolic case, where singularities propagate along characteristics and in general persist over time.

Another qualitative result that can be deduced from our representation, concerns the large time behavior of the solution (i.e. the behavior in the limit $t \to \infty$). This behavior is directly influenced by the boundary conditions. In particular, it depends on the minimal eigenvalue of the corresponding eigenvalue problem. In our case, all the eigenvalues are strictly positive, and from (2.17) and the uniform convergence in D_ε it follows that

$$\lim_{t\to\infty} u(x,t) = 0 \quad \forall 0 \le x \le L.$$

Hence the temperature along the rod converges to the temperature that is imposed at the end points.

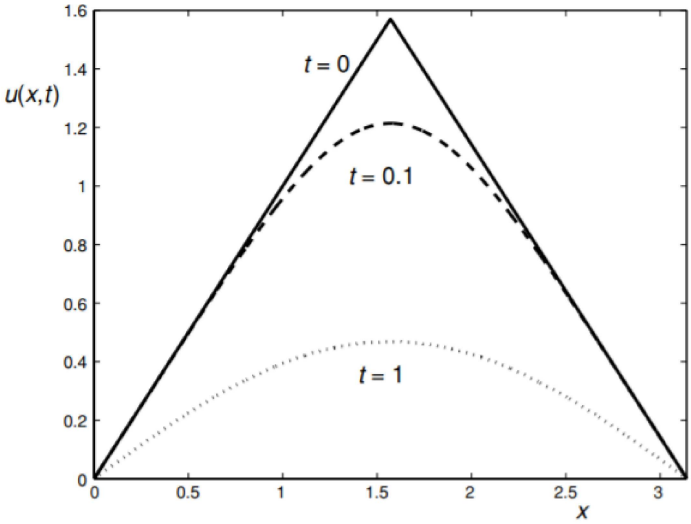

Figure 2.3: The function $u(x,t)$ of (5.27) for $t = 0, t = 0.1$, and $t = 1$. Notice that the singularity at $t = 0$ is quickly smoothed out. The graphs were generated with 200 terms in the Fourier expansion. Actually just three or four terms are needed to capture u correctly even for $t = 0.1$.

We conclude this section by mentioning other boundary conditions that appear frequently in heat conduction problems (Module 1). Specifically, we distinguish between two types of boundary conditions:

(a) Separated boundary conditions These boundary conditions can be written as $B_0[u] = \alpha u(0,t) + \beta u_x(0,t) = 0, \quad B_L[u] = \gamma u(L,t) + \delta u_x(L,t) = 0 \quad t \geq 0$, where

$$\alpha, \beta, \gamma, \delta \in \mathbb{R}, \quad |\alpha| + |\beta| > 0, \quad |\gamma| + |\delta| > 0.$$

This type of boundary condition includes for $\alpha = \gamma = 1, \beta = \delta = 0$ the

Dirichlet boundary condition

$$u(0,t) = u(L,t) = 0 \quad t \geq 0,$$

which is also called a boundary condition of the first kind. Also, for $\alpha = \gamma = 0, \beta = \delta = 1$ we obtain

$$u_x(0,t) = u_x(L,t) = 0 \quad t \geq 0,$$

which is called the Neumann condition or a boundary condition of the second kind. Recall that the physical interpretation of the Neumann boundary condition for heat problems is that there is no heat flow through the boundary. In our case it means that the rod is insulated. If we impose a Dirichlet condition at one end, and a Neumann condition at the other hand, then the boundary condition is called mixed. In the general case, where $\alpha, \beta, \gamma, \delta$ are nonzero, the boundary condition is called a boundary condition of the third kind (or the Robin condition). The physical interpretation is that the heat flow at the boundary depends linearly on the temperature.

(b) Periodic boundary condition This boundary condition is imposed for example in the case of heat evolution along a circular wire of length L. Clearly, in this case the temperature $u(x,t)$ and all its derivatives are periodic (as a function of x) with a period L. In addition u satisfies the heat equation on $(0,L)$. The boundary conditions for this problem are

$$u(0,t) = u(L,t), \quad u_x(0,t) = u_x(L,t) \quad \forall t \geq 0.$$

The periodicity of all the higher-order derivatives follows from the PDE and the boundary conditions presented above.

2.1.3 Separation of variables for the wave equation

We now apply the method of separation of variables to solve the problem of a vibrating string without external forces and with two clamped but free ends. Let $u(x,t)$ be the amplitude of the string at the point x and time t, and let f and g be the amplitude and the velocity of the string at time $t = 0$. We need to solve the problem

$$u_{tt} - c^2 u_{xx} = 0 \quad 0 < x < L, \quad t > 0, \tag{2.29}$$

$$u_x(0,t) = u_x(L,t) = 0 \quad t \geq 0, \tag{2.30}$$

$$u(x,0) = f(x) \quad 0 \leq x \leq L, \tag{2.31}$$

$$u_t(x,0) = g(x) \quad 0 \leq x \leq L, \tag{2.32}$$

where f, g are given functions and c is a positive constant. The compatibility conditions are given by

$$f'(0) = f'(L) = g'(0) = g'(L) = 0.$$

The problem (2.29)-(2.32) is a linear homogeneous initial boundary value problem. As mentioned above, the conditions (5.29) are called Neumann boundary conditions.

Recall that at the first stage of the method, we compute nontrivial separated solutions of the PDE (2.29), i.e. solutions of the form

$$u(x,t) = X(x)T(t),$$

that also satisfy the boundary conditions (2.30). Here, as usual, X, T are functions of the variables x and t respectively. At this stage, we do

not take into account the initial conditions (2.31)-(2.32).

Differentiating the separated solution (2.33) twice in x and twice in t, and then substituting these derivatives into the wave equation, we infer

$$XT_{tt} = c^2 X_{xx} T. \tag{2.33}$$

By separating the variables, we see that

$$\frac{T_{tt}}{c^2 T} = \frac{X_{xx}}{X}. \tag{2.34}$$

It follows that there exists a constant λ such that

$$\frac{T_{tt}}{c^2 T} = \frac{X_{xx}}{X} = -\lambda. \tag{2.35}$$

Equation (2.35) implies

$$\frac{\mathrm{d}^2 X}{\mathrm{d}x^2} = -\lambda X \quad 0 < x < L, \tag{2.36}$$

$$\frac{\mathrm{d}^2 T}{\mathrm{d}t^2} = -\lambda c^2 T \quad t > 0. \tag{2.37}$$

The boundary conditions (2.30) for u imply

$$u_x(0,t) = \frac{\mathrm{d}X}{\mathrm{d}x}(0)T(t) = 0, \quad u_x(L,t) = \frac{\mathrm{d}X}{\mathrm{d}x}(L)T(t) = 0.$$

Since u is nontrivial it follows that

$$\frac{\mathrm{d}X}{\mathrm{d}x}(0) = \frac{\mathrm{d}X}{\mathrm{d}x}(L) = 0.$$

Therefore, the function X should be a solution of the eigenvalue problem

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2} + \lambda X = 0 \quad 0 < x < L, \tag{2.38}$$

$$\frac{\mathrm{d}X}{\mathrm{d}x}(0) = \frac{\mathrm{d}X}{\mathrm{d}x}(L) = 0. \tag{2.39}$$

This eigenvalue problem is also called the Neumann problem.

We have already written the general solution of the ODE (2.38):

1. if $\lambda < 0$, then $X(x) = \alpha\cosh(\sqrt{-\lambda}x) + \beta\sinh(\sqrt{-\lambda}x)$,
2. if $\lambda = 0$, then $X(x) = \alpha + \beta x$,
3. if $\lambda > 0$, then $X(x) = \alpha\cos(\sqrt{\lambda}x) + \beta\sin(\sqrt{\lambda}x)$,

where α, β are arbitrary real numbers.

Negative eigenvalue ($\lambda < 0$) The first boundary condition $(\mathrm{d}X/\mathrm{d}x)(0) = 0$ implies that $\beta = 0$. Then $(\mathrm{d}X/\mathrm{d}x)(L) = 0$ implies that $\sinh(\sqrt{-\lambda}L) = 0$. Therefore, $X(x) \equiv 0$ and the eigenvalue problem (2.38)-(2.39) does not admit negative eigenvalues.

Zero eigenvalue ($\lambda = 0$) The general solution is a linear function $X(x) = \alpha + \beta x$. Substituting this solution into the boundary conditions (2.39) implies that $\lambda = 0$ is an eigenvalue with a unique eigenfunction $X_0(x) \equiv 1$ (the eigenfunction is unique up to a multiplicative factor).

Positive eigenvalue ($\lambda > 0$) The general solution for $\lambda > 0$ has the form

$$X(x) = \alpha\cos(\sqrt{\lambda}x) + \beta\sin(\sqrt{\lambda}x). \tag{2.40}$$

Substituting it in $(\mathrm{d}X/\mathrm{d}x)(0) = 0$, we obtain $\beta = 0$. The boundary condition $(\mathrm{d}X/\mathrm{d}x)(L) = 0$ implies now that $\sin(\sqrt{\lambda}L) = 0$. Thus $\sqrt{\lambda}L = n\pi$, where $n \in \mathbb{N}$. Consequently, $\lambda > 0$ is an eigenvalue if and only if:

$$\lambda = \left(\frac{n\pi}{L}\right)^2 \quad n = 1, 2, 3, \ldots$$

The associated eigenfunction is

$$X(x) = \cos\frac{n\pi x}{L}$$

and it is uniquely determined up to a multiplicative factor.

Therefore, the solution of the eigenvalue problem (2.38)-(2.39) is an infinite sequence of nonnegative simple eigenvalues and their associated eigenfunctions. We use the convenient notation:

$$X_n(x) = \cos\frac{n\pi x}{L}, \quad \lambda_n = \left(\frac{n\pi}{L}\right)^2 \quad n = 0, 1, 2, \ldots$$

Consider now the ODE(2.37) for $\lambda = \lambda_n$. The solutions are

$$T_0(t) = \gamma_0 + \delta_0 t \tag{2.41}$$

$$T_n(t) = \gamma_n \cos\left(\sqrt{\lambda_n c^2}t\right) + \delta_n \sin\left(\sqrt{\lambda_n c^2}t\right) \quad n = 1, 2, 3, \ldots \tag{2.42}$$

Thus, the product solutions of the initial boundary value problem are given by

$$u_0(x,t) = X_0(x)T_0(t) = \frac{A_0 + B_0 t}{2}, \tag{2.43}$$

$$\begin{aligned} u_n(x,t) &= X_n(x)T_n(t) \\ &= \cos\frac{n\pi x}{L}\left(A_n\cos\frac{c\pi nt}{L} + B_n\sin\frac{c\pi nt}{L}\right) \end{aligned} \tag{2.44}$$

For $n = 1, 2, \ldots$

Applying the (generalized) superposition principle, the expression

$$\begin{aligned} &u(x,t) \\ &= \frac{A_0 + B_0 t}{2} + \sum_{n=1}^{\infty}\left(A_n\cos\frac{c\pi nt}{L} + B_n\sin\frac{c\pi nt}{L}\right)\cos\frac{n\pi x}{L} \end{aligned} \tag{2.45}$$

is a (generalized, or at least formal) solution of the problem (2.29)-(2.32). We can show that the solution (2.44) can be represented as a superposition of forward and backward waves. In other words, solution (2.45) is also a generalized solution of the wave equation.

It remains to find the coefficients A_n, B_n in solution (2.45). Here we use the initial conditions. Assume that the initial data f, g can be expanded into generalized Fourier series with respect to the sequence of the eigenfunctions of the problem, and that these series are uniformly converging. That is,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n\cos\frac{n\pi x}{L} \tag{2.46}$$

$$g(x) = \frac{\tilde{a}_0}{2} + \sum_{n=1}^{\infty} \tilde{a}_n\cos\frac{n\pi x}{L} \tag{2.47}$$

Again, the (generalized) Fourier coefficients of f and g can easily be de-

termined; for $m \geq 0$, we multiply (2.46) by the eigenfunction $\cos(m\pi x/L)$, and then we integrate over $[0, L]$. We obtain

$$\int_0^L \cos\frac{m\pi x}{L} f(x)\mathrm{d}x = \frac{a_0}{2}\int_0^L \cos\frac{m\pi x}{L}\ \mathrm{d}x + \sum_{n=1}^{\infty} a_n \int_0^L \cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\ \mathrm{d}x. \tag{2.48}$$

It is easily checked that

$$\int_0^L \cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\ \mathrm{d}x = \begin{cases} 0 & m \neq n, \\ L/2 & m = n \neq 0, \\ L & m = n = 0. \end{cases} \tag{2.49}$$

Therefore, the Fourier coefficients of f with respect to the system of eigenfunctions are

$$a_0 = 2\frac{\int_0^L f(x)\mathrm{d}x}{\int_0^L 1\ \mathrm{d}x} = \frac{2}{L}\int_0^L f(x)\mathrm{d}x \tag{2.50}$$

$$a_m = \frac{\int_0^L \cos(m\pi x/L) f(x)\mathrm{d}x}{\int_0^L \cos^2(m\pi x/L)\mathrm{d}x} = \frac{2}{L}\int_0^L \cos\frac{m\pi x}{L} f(x)\mathrm{d}x \tag{2.51}$$

For $m = 1, 2, \ldots$

The Fourier coefficients $\tilde{a}_n$ of g can be computed similarly. Substituting $t = 0$ into (2.45), and assuming that the corresponding series

converges uniformly, we obtain

$$u(x,0) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos\frac{n\pi x}{L}$$
$$= f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{L}$$

Recall that the (generalized) Fourier coefficients are uniquely determined, and hence $A_n = a_n$ for all $n \geq 0$. In order to compute B_n, we differentiate (2.45) formally (term-by-term) with respect to t and then substitute $t = 0$. We have

$$u_t(x,0) = \frac{B_0}{2} + \sum_{n=1}^{\infty} B_n \frac{c\pi n}{L} \cos\frac{n\pi x}{L}$$
$$= g(x) = \frac{\tilde{a}_0}{2} + \sum_{n=1}^{\infty} \tilde{a}_n \cos\frac{n\pi x}{L}$$

Therefore, $B_n = \tilde{a}_n L/c\pi n$ for all $n \geq 1$. Similarly, $B_0 = \tilde{a}_0$. Thus, the problem is formally solved. The uniqueness issue will be discussed at the end of this chapter.

There is a significant difference between the solution (2.17) of the heat problem and the formal solution (2.45). Each term of the solution (2.17) of the heat equation has a decaying exponential factor which is responsible for the smoothing effect for $t > 0$. In (2.45) we have instead a (nondecaying) trigonometric factor. This is related to the fact that hyperbolic equations preserve the singularities of the given data since the rate of the decay of the generalized Fourier coefficients to zero usually depends on the smoothness of the given function (under the assumption that this function satisfies the prescribed boundary conditions). The precise decay rate of the Fourier coefficients is provided for the classical

Fourier system by the general theory of Fourier analysis .

Example 2.2 Solve the problem

$$\begin{aligned} u_{tt} - 4u_{xx} &= 0 \quad 0 < x < 1, t > 0, \\ u_x(0,t) = u_x(1,t) &= 0 \quad t \geq 0, \\ u(x,0) = f(x) &= \cos^2 \pi x \quad 0 \leq x \leq 1, \\ u_t(x,0) = g(x) &= \sin^2 \pi x \cos \pi x \quad 0 \leq x \leq 1. \end{aligned} \tag{2.52}$$

The solution of (2.52) was shown to have the form

$$\begin{aligned} &u(x,t) \\ &= \frac{A_0 + B_0 t}{2} + \sum_{n=1}^{\infty} (A_n \cos 2n\pi t + B_n \sin 2n\pi t) \cos n\pi x. \end{aligned} \tag{2.53}$$

Substituting f into (2.53) implies

$$u(x,0) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos n\pi x = \cos^2 \pi x. \tag{2.54}$$

The Fourier expansion of f is easily obtained using the trigonometric identity $\cos^2 \pi x = \frac{1}{2} + \frac{1}{2} \cos 2\pi x$. Since the Fourier coefficients are uniquely determined, it follows that

$$A_0 = 1, A_2 = \frac{1}{2}, A_n = 0 \quad \forall n \neq 0, 2. \tag{2.55}$$

By differentiating the solution with respect to t, and substituting $u_t(x,0)$ into the second initial condition, we obtain

$$u_t(x,0) = \frac{B_0}{2} + \sum_{n=1}^{\infty} B_n 2n\pi \cos n\pi x = \sin^2 \pi x \cos \pi x \tag{2.56}$$

Similarly, the Fourier expansion of g is obtained using the trigonometric identity $\sin^2 \pi x \cos \pi x = \frac{1}{4}\cos \pi x - \frac{1}{4}\cos 3\pi x$. From the uniqueness of the expansion it follows that

$$B_1 = \frac{1}{8\pi}, B_3 = -\frac{1}{24\pi}, B_n = 0 \quad \forall n \neq 1, 3.$$

Therefore,

$$u(x,t) = \frac{1}{2} + \frac{1}{8\pi}\sin 2\pi t \cos \pi x + \frac{1}{2}\cos 4\pi t \cos 2\pi x - \frac{1}{24\pi}\sin 6\pi t \cos 3\pi x. \tag{2.57}$$

Since (2.57) contains only a finite number of (smooth) terms, it is verified directly that u is a classical solution of the problem.

2.1.4 Separation of variables for nonhomogeneous equations

It is possible to upgrade the method of separation of variables to a method for solving nonhomogeneous PDEs. This technique is called also the method of eigenfunction expansion. For example, consider the problem

$$\begin{aligned} u_{tt} - u_{xx} &= \cos 2\pi x \cos 2\pi t \qquad 0 < x < 1, t > 0, \\ u_x(0,t) &= u_x(1,t) = 0 \qquad t \geq 0, \\ u(x,0) &= f(x) = \cos^2 \pi x \qquad 0 \leq x \leq 1, \\ u_t(x,0) &= g(x) = 2\cos 2\pi x \qquad 0 \leq x \leq 1. \end{aligned} \tag{2.58}$$

In the previous section we found the system of all eigenfunctions and the corresponding eigenvalues of the homogeneous problem. They are

$$X_n(x) = \cos n\pi x, \quad \lambda_n = (n\pi)^2 \quad n = 0, 1, 2, \ldots.$$

Recall Fourier's claim that any reasonable function satisfying the boundary conditions can be uniquely expanded into (generalized) Fourier series with respect to the system of the eigenfunctions of the problem. Since the solution $u(x,t)$ of the problem (2.58) is a twice differentiable function satisfying the boundary conditions, it follows that for a fixed t the solution u can be represented as

$$u(x,t) = \frac{1}{2}T_0(t) + \sum_{n=1}^{\infty} T_n(t) \cos n\pi x \tag{2.59}$$

where $T_n(t)$ are the (time dependent) Fourier coefficients of the function $u(\cdot, t)$. Hence, we need to find these coefficients.

Substituting (2.59) into the wave equation (2.58) and differentiating the series term-by-term implies that

$$\frac{1}{2}T_0'' + \sum_{n=1}^{\infty} \left(T_n'' + n^2\pi^2 T_n\right) \cos n\pi x = \cos 2\pi t \cos 2\pi x \tag{2.60}$$

Note that in the current example, the right hand side of the equation

is already given in the form of a Fourier series. The uniqueness of the Fourier expansion implies that the Fourier coefficients of the series of the left hand side of (2.60) are equal to the Fourier coefficients of the series of the right hand side. In particular, for $n = 0$ we obtain the ODE:

$$T_0'' = 0 \tag{2.61}$$

whose general solution is $T_0(t) = A_0 + B_0 t$. Similarly we obtain for $n = 2$

$$T_2'' + 4\pi^2 T_2 = \cos 2\pi t \tag{2.62}$$

The general solution of this linear nonhomogeneous second-order ODE is

$$T_2(t) = A_2 \cos 2\pi t + B_2 \sin 2\pi t + \frac{t}{4\pi} \sin 2\pi t.$$

For $n \neq 0, 2$, we have

$$T_n'' + n^2\pi^2 T_n = 0 \quad \forall n \neq 0, 2 \tag{2.63}$$

The solution is $T_n(t) = A_n \cos n\pi t + B_n \sin n\pi t$ for all $n \neq 0, 2$. Substituting the solutions of (2.61), (2.62), and (2.63) into (2.59) implies that the solution of the problem is of the form

$$\begin{aligned} u(x,t) &= \frac{A_0 + B_0 t}{2} + \frac{t}{4\pi} \sin 2\pi t \cos 2\pi x \\ &\quad + \sum_{n=1}^{\infty} (A_n \cos n\pi t + B_n \sin n\pi t) \cos n\pi x. \end{aligned} \tag{2.64}$$

Substituting (2.64) into the first initial condition (2.58), we get

$$u(x,0) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos n\pi x = \cos^2 \pi x = \frac{1}{2} + \frac{1}{2}\cos 2\pi x,$$

therefore,

$$A_0 = 1, \quad A_2 = \frac{1}{2}, \quad A_n = 0 \forall n \neq 0, 2.$$

By differentiating (term-by-term) the solution u with respect to t and substituting $u_t(x,0)$ into the second initial condition of (2.58), we find

$$u_t(x,0) = \frac{B_0}{2} + \sum_{n=1}^{\infty} n\pi B_n \cos n\pi x = 2\cos 2\pi x,$$

Hence,

$$B_2 = \frac{1}{\pi}, \quad B_n = 0 \forall n \neq 2.$$

Finally

$$u(x,t) = \frac{1}{2} + \left(\frac{1}{2}\cos 2\pi t + \frac{t+4}{4\pi}\sin 2\pi t\right)\cos 2\pi x.$$

It is clear that this solution is classical, since the (generalized) Fourier series of the solution has only a finite number of nonzero smooth terms, and therefore all the formal operations are justified. Note that the amplitude of the vibrating string grows linearly in t and it is unbounded as $t \to \infty$.

2.1.5 The energy method and uniqueness

The energy method is a fundamental tool in the theory of PDEs. One of its main applications is in proving the uniqueness of the solution of initial

boundary value problems. The method is based on the physical principle of energy conservation, although in some applications the object we refer to mathematically as an 'energy' is not necessarily the actual energy of a physical system. Recall that in order to prove the uniqueness of solutions for a linear differential problem, it is enough to show that the solution of the corresponding homogeneous PDE with homogeneous initial and boundary conditions is necessarily the zero solution. This basic principle has already been used in module 1 and will be demonstrated again below.

Let us outline the energy method. For certain homogeneous problems it is possible to define an energy integral that is nonnegative and is a nonincreasing function of the time t. In addition, for $t = 0$ the energy is zero and therefore, the energy is zero for all $t \geq 0$. Due to the positivity of the energy, and the zero initial and boundary conditions it will follow that the solution is zero.

We demonstrate the energy method for the problems that have been studied in the present chapter.

Example 2.3 Consider the Neumann problem for the vibrating string

$$u_{tt} - c^2 u_{xx} = F(x,t) \quad 0 < x < L, t > 0, \tag{2.65}$$

$$u_x(0,t) = a(t), \quad u_x(L,t) = b(t) \quad t \geq 0, \tag{2.66}$$

$$u(x,0) = f(x) \quad 0 \leq x \leq L, \tag{2.67}$$

$$u_t(x,0) = g(x) \quad 0 \leq x \leq L. \tag{2.68}$$

Let u_1, u_2 be two solutions of the problem. By the superposition principle, the function $w := u_1 - u_2$ is a solution of the problem

$$w_{tt} - c^2 w_{xx} = 0 \quad 0 < x < L, t > 0, \tag{2.69}$$

$$w_x(0,t) = 0, w_x(L,t) = 0 \quad t \geq 0, \tag{2.70}$$

$$w(x,0) = 0 \quad 0 \leq x \leq L, \tag{2.71}$$

$$w_t(x,0) = 0 \quad 0 \leq x \leq L. \tag{2.72}$$

Define the total energy of the solution w at time t as

$$E(t) := \frac{1}{2}\int_0^L \left(w_t^2 + c^2 w_x^2\right) \mathrm{d}x. \tag{2.73}$$

The first term represents the total kinetic energy of the string, while the second term is the total potential energy. Clearly, E' is given by

$$\begin{aligned} E'(t) & \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}\int_0^L \left(w_t^2 + c^2 w_x^2\right) \mathrm{d}x\right] \\ &= \int_0^L \left(w_t w_{tt} + c^2 w_x w_{xt}\right) \mathrm{d}x. \end{aligned} \tag{2.74}$$

But

$$c^2 w_x w_{xt} = c^2\left[\frac{\partial}{\partial x}\left(w_x w_t\right) - w_{xx} w_t\right] = c^2 \frac{\partial}{\partial x}\left(w_x w_t\right) - w_{tt} w_t$$

Substituting this identity into (2.74) and using the fundamental theorem of calculus, we have

$$E'(t) = c^2 \int_0^L \frac{\partial}{\partial x}\left(w_x w_t\right) \mathrm{d}x = c^2 \left(w_x w_t\right)\Big|_0^L.$$

The boundary condition (2.70) implies that $E'(t) = 0$, hence, $E(t) =$ constant and the energy is conserved.

On the other hand, since for $t = 0$ we have $w(x, 0) = 0$, it follows that $w_x(x, 0) = 0$. Moreover, we have also $w_t(x, 0) = 0$. Therefore, the energy at time $t = 0$ is zero. Thus, $E(t) \equiv 0$.

Since $e(x, t) := w_t^2 + c^2 w_x^2 \geq 0$, and since its integral over $[0, L]$ is zero, it follows that $w_t^2 + c^2 w_x^2 \equiv 0$, which implies that $w_t(x, t) = w_x(x, t) \equiv 0$. Consequently, $w(x, t) \equiv$ constant. By the initial conditions $w(x, 0) = 0$, hence $w(x, t) \equiv 0$. This completes the proof of the uniqueness of the problem (2.65)-(2.68).

Example 2.4 Let us modify the previous problem a little, and instead of the (nonhomogeneous) Neumann problem, consider the Dirichlet boundary conditions:

$$u(0, t) = a(t), \quad u(L, t) = b(t) \quad t \geq 0.$$

We use the same energy integral and follow the same steps. We obtain for the function w

$$E'(t) = c^2 \left(w_x w_t\right)\Big|_0^L . \tag{2.75}$$

Since $w(0, t) = w(L, t) = 0$, it follows that $w_t(0, t) = w_t(L, t) = 0$; therefore, $E'(t) = 0$ and in this case too the energy is conserved. The rest of the proof is exactly the same as in the previous example.

Example 2.5 The energy method can also be applied to heat conduction problems. Consider the Dirichlet problem

$$u_t - k u_{xx} = F(x, t) \quad 0 < x < L, t > 0, \tag{2.76}$$

$$u(0, t) = a(t), u(L, t) = b(t) \quad t \geq 0, \tag{2.77}$$

$$u(x, 0) = f(x) \quad 0 \leq x \leq L. \tag{2.78}$$

As we explained above, we need to prove that if w is a solution of the homogeneous problem with zero initial and boundary conditions, then $w = 0$. In the present case, we define the energy to be:

$$E(t) := \frac{1}{2}\int_0^L w^2 \, \mathrm{d}x \tag{2.79}$$

The time derivative E' is given by

$$E'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\int_0^L w^2 dx\right) = \int_0^L w w_t \, \mathrm{d}x = \int_0^L k w w_{xx} \, \mathrm{d}x. \tag{2.80}$$

Integrating by parts and substituting the boundary conditions, we have

$$E'(t) = k w w_x|_0^L - \int_0^L k \left(w_x\right)^2 \, \mathrm{d}x = -\int_0^L k \left(w_x\right)^2 \, \mathrm{d}x \leq 0$$

therefore, the energy is not increasing. Since $E(0) = 0$ and $E(t) \geq 0$, it follows that $E \equiv 0$. Consequently, for all $t \geq 0$ we have $w(\cdot, t) \equiv 0$ and the uniqueness is proved. The same proof can also be used for the Neumann problem and even for the boundary condition of the third kind:

$$u(0,t) - \alpha u_x(0,t) = a(t). \quad u(L,t) + \beta u_x(L,t) = b(t) \quad t \geq 0,$$

provided that $\alpha, \beta \geq 0$.

2.1.6 Further applications of the heat equation

We have seen that the underlying property of the wave equation is to propagate waves, while the heat equation smoothes out oscillations and discontinuities. In this section we shall consider two specific applications

of the heat equation that concern signal propagation. In the first application we shall show that a diffusion mechanism can still transmit (to some extent) oscillatory data. In fact, diffusion effects play an important role in one of the most important communication systems. In the second example the goal will be to use the smoothing property of the heat equation to dampen oscillations in the data.

The cable equation

The great success of the telegraph prompted businessmen and governments to lay an underwater cable between France and Britain in 1850 . It was realized, however, that the transmission rate through this cable was very low. The British scientist William Thomson (1824-1907) sought to explain this phenomenon. His mathematical model showed that the cable's electrical capacity has a major effect on signal transmission. We shall derive the equation for signal transmission in a cable, solve it, and then explain Thomson's analysis.

A cross section of the cable is shown in Figure 2.4. The cable is modeled as a system of outer and inner conductors separated by an insulating layer. To simplify the analysis we shall consider a two-dimensional model, using x to denote the

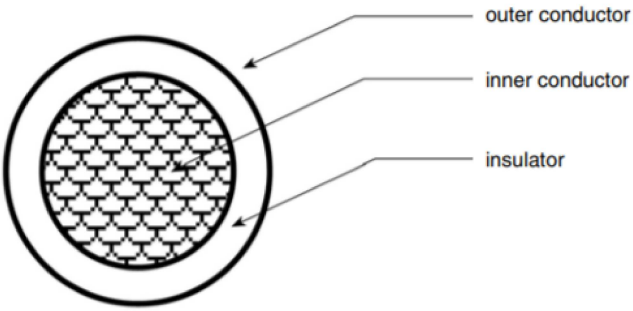

Figure 2.4: Cross section of cable

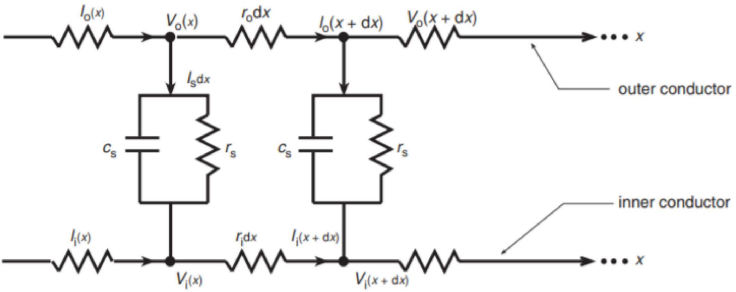

Figure 2.5: A longitudinal cross section of the cable.

Figure 2.5 A longitudinal cross section of the cable.

longitudinal direction. A small segment of the longitudinal cross section is shown in Figure 2.5. In this segment we see the local resistivity (r_{i} in the inner conductor, and r_{o} in the outer conductor), while the insulator is modeled by a capacitor C_{s} and a resistor r_{s} in parallel. The transversal current in a horizontal element of length $\mathrm{d}x$ is $I_{\mathrm{s}}\mathrm{d}x$. Ohm's

law for the segment $(x, x + \mathrm{d}x)$ implies

$$V_{\mathrm{i}}(x+\mathrm{d}x)-V_{\mathrm{i}}(x) = -I_{\mathrm{i}}(x)r_{\mathrm{i}}\mathrm{d}x, V_{\mathrm{o}}(x+\mathrm{d}x)-V_{\mathrm{o}}(x) = -I_{\mathrm{o}}(x)r_{\mathrm{o}}\mathrm{d}x. \quad (2.81)$$

In the limit $\mathrm{d}x \to 0$ this becomes

$$\frac{\partial V_{\mathrm{i}}}{\partial x} = -r_{\mathrm{i}}I_{\mathrm{i}}(x), \quad \frac{\partial V_{\mathrm{o}}}{\partial x} = -r_{\mathrm{o}}I_{\mathrm{o}}(x). \quad (2.82)$$

Having balanced the voltage drop in the longitudinal direction, we proceed to write the current conservation equation (Kirchhoff's law). We have

$$I_{\mathrm{i}}(x + \mathrm{d}x) = I_{\mathrm{i}}(x) + I_{\mathrm{s}}(x)\mathrm{d}x, \quad I_{\mathrm{o}}(x + \mathrm{d}x) = I_{\mathrm{o}}(x) - I_{\mathrm{s}}(x)\mathrm{d}x. \quad (2.83)$$

Having balanced the voltage drop in the longitudinal direction, we proceed to write the current conservation equation (Kirchhoff's law). We have

$$I_{\mathrm{i}}(x + \mathrm{d}x) = I_{\mathrm{i}}(x) + I_{\mathrm{s}}(x)\mathrm{d}x, \quad I_{\mathrm{o}}(x + \mathrm{d}x) = I_{\mathrm{o}}(x) - I_{\mathrm{s}}(x)\mathrm{d}x. \quad (2.84)$$

Again letting $\mathrm{d}x \to 0$ we obtain

$$I_{\mathrm{s}} = \frac{\partial I_{\mathrm{i}}}{\partial x} = -\frac{\partial I_{\mathrm{o}}}{\partial x}. \quad (2.85)$$

Introducing the transinsulator potential $V = V_{\mathrm{i}} - V_{\mathrm{o}}$, we conclude from (2.83) that

$$-\frac{\partial V}{\partial x} = r_{\mathrm{i}}I_{\mathrm{i}} - r_{\mathrm{o}}I_{\mathrm{o}} \quad (2.86)$$

Differentiating (2.86) by x, and using (2.85) we get

$$-I_{\mathrm{s}} = \frac{1}{r_{\mathrm{i}} + r_{\mathrm{o}}} \frac{\partial^2 V}{\partial x^2} \tag{2.87}$$

It remains to understand the current I_{s}. The contribution of the resistor r_{s} is $-(1/r_{\mathrm{s}})\, V$. The current through a capacitor is given by $-C_{\mathrm{s}} \partial V/\partial t$, where C_{s} denotes the capacitance. Therefore we finally obtain the (passive) cable equation

$$\frac{\partial V}{\partial t} = D \frac{\partial^2 V}{\partial x^2} - \beta V \quad D = \frac{1}{C_{\mathrm{s}}\,(r_{\mathrm{i}} + r_{\mathrm{e}})}, \quad \beta = \frac{1}{r_{\mathrm{s}} C_{\mathrm{s}}}. \tag{2.88}$$

Note that the capacitor gave rise to a diffusion-like term in the transport equation. Equation (2.88) can be solved in a finite x interval by the separation of variables method In order to understand its use in communication, we shall assume that the transmitter is located at $x = 0$, and the receiver is at an arbitrary location x up the line. Therefore we solve the cable equation for a semi-infinite interval. To fix ideas, we formulate the following problem:

$$V_t = D V_{xx} - \beta V \quad 0 < x < \infty, \quad -\infty < t < \infty, \tag{2.89}$$

$$V(0,t) = A \cos \omega t \quad -\infty < t < \infty, \tag{2.90}$$

$$V(x,t) \to 0 \quad x \to \infty. \tag{2.91}$$

The problem (2.89)-(2.91) can be solved by a variant of the separation of variables method. Our motivation is to seek a solution that will have propagation and oscillation properties as in a wave equation, but also

decay properties that are typical of a heat equation. Therefore we seek a solution of the form

$$V(x,t) = Av(x)\cos(\omega t - kx). \tag{2.92}$$

Substituting (2.92) into (2.89), and defining $\phi := \omega t - kx$, we get

$$-\omega v \sin\phi = D\left(v_{xx}\cos\phi + 2kv_x\sin\phi - k^2 v\cos\phi\right) - \beta v\cos\phi.$$

We first equate the coefficients of the $\cos\phi$ term. This implies

$$v_{xx} - \left(k^2 + \frac{\beta}{D}\right)v = 0.$$

The boundary conditions (2.90) − (2.91) imply $v(0) = 1, v(\infty) = 0$. Therefore $v(x) = \exp\left[-\sqrt{k^2 + (\beta/D)}x\right]$. Equating now the coefficients of the term $\sin\phi$, using the solution that was found for v, we find that (2.92) is indeed the desired solution if k, ω, and D satisfy the dispersion relation

$$\omega = 2Dk\sqrt{k^2 + \frac{\beta}{D}}. \tag{2.93}$$

We now analyze this solution in light of the cable transmission issue. The parameter β represents the loss of energy due to the transinsulator resistivity. Increasing the resistivity will decrease β. We therefore proceed to consider an ideal situation where $\beta = 0$. In this case the solution (2.92) and the dispersion relation (2.93) become

$$V(x,t) = Ae^{-kx}\cos(\omega t - kx), \quad \omega = 2Dk^2. \tag{2.94}$$

The frequency ω can be interpreted as the rate of transmission. Sim-

ilarly, we interpret $1/k$ as the distance L between the transmitter and the receiver. Therefore $\omega = 2DL^{-2}$. This formula enabled Thomson to predict that with the parameters of the materials used for the cable, i.e. C_s, r_i, r_o that determine D, and in light of the distance L, the transmission rate would be far below the expected rate. His prediction was indeed fulfilled. Following the great success of his mathematical analysis, Thomson was asked to consult in the next major attempt to lay an underwater communication cable, this time in 1865 between Britain and the USA. The great improvement in production control allowed the manufacture of a high quality cable, and the enterprise met with high technical and financial success. To honor him for his contributions to the transatlantic cable, Thomson was created Lord Kelvin in 1866.

Interest in the cable equation was renewed in the middle of the twentieth century when it was discovered to be an adequate model for signal transmission in biological cells in general, and for neurons in particular. The insulating layer in this case is the cell's membrane. The currents consist of ions, mainly potassium and sodium ions.

In the biological applications, however, one needs to replace the passive resistor r_s with a nonlinear electrical element. The reason is that the current through the cell's membrane flows in special channels with a complex gate mechanism that was deciphered by Hodgkin and Huxley . Moreover, we need to supplement in this case the resulting active cable equation with further dynamical rules for the channel gates.

As another example of the energy method we shall prove that the solution we found for the cable equation is unique. More precisely, we shall prove that the problem (2.89)-(2.91) has a unique solution in the class of functions for which the energy

$$E_w(t) = E(t) := \frac{1}{2}\int_0^\infty [w(x,t)]^2 \, \mathrm{d}x \tag{2.95}$$

is bounded. Namely, we assume that for each solution w of the problem there exists a constant $M_w > 0$ such that $E_w(t) \leq M_w$. We need to prove that if w is a solution of the homogeneous problem with zero boundary conditions, then $w = 0$. We obtain as for the heat equation that

$$\begin{aligned} E'(t) & \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\int_0^\infty w^2 \, \mathrm{d}x\right) \\ &= \int_0^\infty w w_t \, \mathrm{d}x \\ &= \int_0^\infty \left(D w w_{xx} - \beta w^2\right) \mathrm{d}x, \end{aligned} \tag{2.96}$$

where we have assumed that all the above integrals are finite. Integrating by parts and substituting the boundary conditions, we have

$$E'(t) \leq -\int_0^\infty D\left(w_x\right)^2 \, \mathrm{d}x - \beta E(t) \leq -\beta E(t).$$

Fixing $T \in \mathbb{R}$ and integrating the above differential inequality from t to $T > t$, we obtain the estimate

$$E(T) \leq E(t)\mathrm{e}^{-\beta(T-t)} \leq M\mathrm{e}^{-\beta T}\mathrm{e}^{\beta t}.$$

Letting $t \to -\infty$ it follows that $E(T) = 0$. Therefore, $E \equiv 0$ which implies that $w = 0$.

2.1.7 Wine cellars

Most types of foodstuff require good temperature control. A well-known example is wine, which is stored in underground wine cellars. The idea is that a good layer of soil will shield the wine from temperature fluctuations with the seasons (and even daily fluctuations). Clearly very deep cellars will do this, but such cellars are costly to build, and inconvenient to use and maintain. Therefore we shall use the solution we found in the previous section for the heat equation in a semi-infinite strip to estimate an adequate depth for a wine cellar. We consider the following model:

$$u_t = Du_{xx} \quad 0 < x < \infty, -\infty < t < \infty, \tag{2.97}$$

$$u(0,t) = T_0 + A\cos\omega t \quad -\infty < t < \infty, \tag{2.98}$$

$$u(x,t) \to T_0 \quad x \to \infty. \tag{2.99}$$

Here the x coordinate measures the distance towards the earth center, where $x = 0$ is the earth's surface, D is the earth's diffusion coefficient, and ω represents the ground temperature fluctuations about a fixed temperature T_0. For example, one can take one year as the basic period, which implies $\omega = 0.19 \times 10^{-6}$ s^{-1}. Thanks to the superposition principle and formula (2.94), we obtain the solution:

$$V(x,t) = T_0 + Ae^{-kx}\cos(\omega t - kx), \quad \omega = 2Dk^2. \tag{2.100}$$

How should formula (2.100) be used to choose the depth of the cellar? We have already determined ω. The diffusion coefficient D depends on the nature of the soil. It can vary by a factor of 5 or more between dry soil and wet soil and rocks. For the purpose of our model we shall assume an average value of 0.0025 cm^2 s^{-1}. The ground temperature can fluctuate

by 20°C. If we want to minimize the fluctuation in the cellar to less than 2°C, say, we need to use a depth L such that $e^{-kL} = 0.1$, i.e. $L = 3.7$ m. A smarter choice for the depth L would be the criterion $kL = \pi$, i.e. $L = 5$ m. This will provide two advantages. First, it gives a reduction in the amplitude by a factor of 23 , i.e. the fluctuation will be less than 1°C. Second, the phase at this depth would be exactly opposite to the phase at zero ground level (with respect to the fixed temperature T_0). This effect is desirable, since other mechanisms of heat transfer, such as opening the door to the cellar, convection of heat by water, etc. would then drive the temperature in the cellar further towards T_0.

Exercises

Problem 2.1.1. *Solve the equation*

$$u_t = 17u_{xx} \quad 0 < x < \pi, t > 0$$

with the boundary conditions

$$u(0,t) = u(\pi,t) = 0 \quad t \geq 0$$

and the initial conditions

$$u(x,0) = \begin{cases} 0 & 0 \leq x \leq \pi/2 \\ 2 & \pi/2 < x \leq \pi \end{cases}$$

Problem 2.1.2. *Prove that the solution we found by separation of variables for the vibration of a free string can be represented as a superposition of a forward and a backward wave.*

Problem 2.1.3. *(a) Using the separation of variables method find a*

(formal) solution of a vibrating string with fixed ends:

$$\begin{aligned} u_{tt} - c^2 u_{xx} &= 0 && 0 < x < L, 0 < t, \\ u(0,t) = u(L,t) &= 0 && t \geq 0, \\ u(x,0) &= f(x) && 0 \leq x \leq L, \\ u_t(x,0) &= g(x) && 0 \leq x \leq L. \end{aligned}$$

(b) Prove that the above solution can be represented as a superposition of a forward and a backward wave.

Problem 2.1.4. *(a) Find a formal solution of the problem*

$$\begin{aligned} u_{tt} &= u_{xx} && 0 < x < \pi, t > 0, \\ u(0,t) = u(\pi,t) &= 0 && t \geq 0, \\ u(x,0) &= \sin^3 x && 0 \leq x \leq \pi, \\ u_t(x,0) &= \sin 2x && 0 \leq x \leq \pi. \end{aligned}$$

(b) Show that the above solution is classical.

Problem 2.1.5. *(a) Using the method of separation of variables, find a (formal) solution of the problem*

$$\begin{aligned} u_t - k u_{xx} &= 0 && 0 < x < L, t > 0, \\ u_x(0,t) = u_x(L,t) &= 0 && t \geq 0, \\ u(x,0) &= f(x) && 0 \leq x \leq L, \end{aligned}$$

describing the heat evolution of an insulated one-dimensional rod (Neumann problem).

(b) Solve the heat equation $u_t = 12 u_{xx}$ in $0 < x < \pi, t > 0$ subject to the

following boundary and initial conditions:

$$u_x(0,t) = u_x(\pi,t) = 0 \quad t \geq 0,$$
$$u(x,0) = 1 + \sin^3 x \qquad 0 \leq x \leq \pi.$$

(c) Find $\lim_{t\to\infty} u(x,t)$ *for all* $0 < x < \pi$*, and explain the physical interpretation of your result.*

Problem 2.1.6. *(a) Using the separation of variables method find a (formal) solution of the following periodic heat problem:*

$$u_t - ku_{xx} = 0 \qquad 0 < x < 2\pi, t > 0,$$
$$u(0,t) = u(2\pi,t), u_x(0,t) = u_x(2\pi,t) \quad t \geq 0,$$
$$u(x,0) = f(x) \qquad 0 \leq x \leq 2\pi,$$

where f *is a smooth periodic function. This system describes the heat evolution on a circular insulated wire of length* 2π*. (b) Find* $\lim_{t\to\infty} u(x,t)$ *for all* $0 < x < 2\pi$*, and explain the physical interpretation of your result. 126 The method of separation of variables (c) Show that if* v *is an arbitrary partial derivative of the solution* u*, then* $v(0,t) = v(2\pi,t)$ *for all* $t \geq 0$*.*

Problem 2.1.7. *Solve the following heat problem:*

$$u_t - ku_{xx} = A\cos\alpha t \qquad 0 < x < 1, t > 0,$$
$$u_x(0,t) = u_x(1,t) = 0 \quad t \geq 0,$$
$$u(x,0) = 1 + \cos^2 \pi x \quad 0 \leq x \leq 1.$$

Problem 2.1.8. *Consider the problem*

$$\begin{aligned} u_t - u_{xx} &= \mathrm{e}^{-t}\sin 3x \quad && 0 < x < \pi, t > 0, \\ u(0,t) = u(\pi,t) &= 0 && t \geq 0, \\ u(x,0) &= f(x) && 0 \leq x \leq \pi. \end{aligned}$$

(a) Solve the problem using the method of eigenfunction expansion.
(b) Find $u(x,t)$ for $f(x) = x\sin x$.
(c) Show that the solution $u(x,t)$ is indeed a solution of the equation

$$u_t - u_{xx} = \mathrm{e}^{-t}\sin 3x \quad 0 < x < \pi, t > 0.$$

Problem 2.1.9. *Consider the problem*

$$\begin{aligned} u_t - u_{xx} - hu &= 0 && 0 < x < \pi, t > 0 \\ u(0,t) = u(\pi,t) &= 0 && t \geq 0 \\ u(x,0) &= x(\pi - x) && 0 \leq x \leq \pi \end{aligned}$$

where h is a real constant.
(a) Solve the problem using the method of eigenfunction expansion. (b) Does $\lim_{t\to\infty} u(x,t)$ exist for all $0 < x < \pi$?
(Hint: Distinguish between the following cases:
(i) $h < 1$,
(ii) $h = 1$,
(iii) $h > 1$.)

Problem 2.1.10. *Consider the problem*

$$\begin{aligned} u_t &= u_{xx} + \alpha u \qquad 0 < x < 1, t > 0 \\ u(0,t) &= u(1,t) = 0 \quad t \geq 0 \\ u(x,0) &= f(x) \qquad\quad 0 \leq x \leq 1, f \in C([0,1]) \end{aligned}$$

(a) Assume that $\alpha = -1$ *and* $f(x) = x$ *and solve the problem.*
(b) Prove that for all $\alpha \leq 0$ *and all* f*, the solution* u *satisfies* $\lim_{t\to\infty} u(x,t) = 0$*.*
(c) Assume now that $\pi^2 < \alpha < 4\pi^2$*. Does* $\lim_{t\to\infty} u(x,t)$ *exist for all* f *? If your answer is no, find a necessary and sufficient condition on* f *which ensures the existence of this limit.*

2.2 Elliptic equations

2.2.1 Introduction

We mentioned in Module 1 the central role played by the Laplace operator in the theory of PDEs. In this chapter we shall concentrate on elliptic equations, and, in particular, on the main prototype for elliptic equations, which is the Laplace equation itself:

$$\Delta u = 0. \tag{2.101}$$

We start by reviewing a few basic properties of elliptic problems. We then introduce the maximum principle, and also formulate a similar principle for the heat equation. We prove the uniqueness and stability of solutions to the Laplace equation in two ways. One approach is based on the maximum principle, and the other approach uses the method of

Green's identities. The simplest solution method for the Laplace equation is the method of separation of variables. Indeed, this method is only applicable in simple domains, such as rectangles, disks, rings, etc., but these domains are often encountered in applications. Moreover, explicit solutions in simple domains provide an insight into the solution's structure in more general domains. Towards the end of the chapter we shall introduce Poisson's kernel formula.

2.2.2 Basic properties of elliptic problems

We limit the discussion in this chapter to functions $u(x, y)$ in two independent variables, although most of the analysis can be readily generalized to higher dimensions. We further limit the discussion to the case where the equation contains only the principal part, and this part is in a canonical form. Nevertheless, we allow for a nonhomogeneous term in the equation. We denote by D a planar domain (i.e. a nonempty connected and open set in $\mathbb{R}^2$). The Laplace equation is given by

$$\Delta u := u_{xx} + u_{yy} = 0 \quad (x, y) \in D. \tag{2.102}$$

A function u satisfying (2.98) is called a harmonic function. The Laplace equation is a special case of a more general equation:

$$\Delta u = F(x, y), \tag{2.103}$$

where F is a given function. Equation (2.99) was used by the French mathematician Simeon Poisson (1781-1840) in his studies of diverse problems in mechanics, gravitation, electricity, and magnetism. Therefore it is called Poisson's equation. In order to obtain a heuristic understanding

of the results to be derived below, it is useful to provide Poisson's equation with a simple physical interpretation. For this purpose we recall from the discussion in Chapter 1 that the solution of Poisson's equation represents the distribution of temperature u in a domain D at equilibrium. The nonhomogeneous term F describes (up to a change of sign) the rate of heat production in D. For the benefit of readers who are familiar with the theory of electromagnetism, we point out that u could also be interpreted as the electric potential in the presence of a charge density $-F$.

In order to obtain a unique temperature distribution, we must provide conditions for the temperature (or temperature flux) at the boundary ∂D. There are several basic boundary conditions (see the discussion in Module 1).

Definition 2.2.1. *The problem defined by Poisson's equation and the Dirichlet boundary condition*

$$u(x, y) = g(x, y) \quad (x, y) \in \partial D \tag{2.104}$$

for a given function g, is called the Dirichlet problem.

In Figure 2.6 we depict the problem schematically.

Definition 2.2.2. *The problem defined by Poisson's equation and the Neumann boundary condition*

$$\partial_n u(x, y) = g(x, y) \quad (x, y) \in \partial D, \tag{2.105}$$

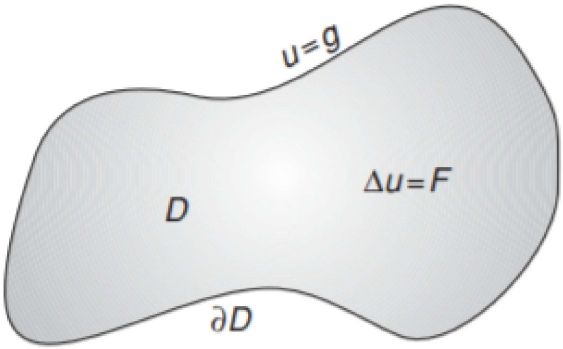

Figure 2.6: A schematic drawing for the Poisson equation with Dirichlet boundary conditions.

Where g is a given function, $\hat{n}$ denotes the unit outward normal to ∂D, and ∂_n denotes a differentiation in the direction of $\hat{n}$ (i.e. $\partial_n = \hat{n}\cdot\nabla$), is called the Neumann problem.

Definition 2.2.3. *The problem defined by Poisson's equation and the boundary condition of the third kind*

$$u(x,y) + \alpha(x,y)\partial_n u(x,y) = g(x,y) \quad (x,y) \in \partial D, \tag{2.106}$$

where α and g are given functions, is called a problem of the third kind (it is also sometimes called the Robin problem).

The first question we have to address is whether there exists a solution to each one of the problems we just defined. This question is not at all easy. It has been considered by many great mathematicians since the middle of the nineteenth century. It was discovered that when the

domain D is bounded and 'sufficiently smooth', then the Dirichlet problem, for example, does indeed have a solution. The precise definition of smoothness in this context and the general existence proof are beyond the scope of this book. It is interesting to point out that in applications one frequently encounters domains with corners (rectangles, for example). Near a corner the boundary is not differentiable; thus, we cannot always expect the solutions to be as smooth as we would like. In this chapter, we only consider classical solutions, i.e. the solutions are in the class $C^2(D)$. Some of the analysis we present requires further conditions on the behavior of the solutions near the boundary. For example, we sometimes have to limit ourselves to solutions in the class $C^1(\bar{D})$.

Consider now the Neumann problem. Since the temperature is in equilibrium, the heat flux through the boundary must be balanced by the temperature production inside the domain. This simple argument is the physical manifestation of the following statement.

Lemma 2.2.1. *A necessary condition for the existence of a solution to the Neumann problem is*

$$\int_{\partial D} g(x(s), y(s))\mathrm{d}s = \int_D F(x, y)\mathrm{d}x\ \mathrm{d}y, \tag{2.107}$$

where $(x(s), y(s))$ *is a parameterization of* ∂D.

Proof Let us first recall the vector identity $\Delta u = \vec{\nabla} \cdot \vec{\nabla} u$. Therefore we can write Poisson's equation as

$$\vec{\nabla} \cdot \vec{\nabla} u = F \tag{2.108}$$

Integrating both sides of the equation over D, and using Gauss' theorem,

we obtain

$$\int_{\partial D} \vec{\nabla} u \cdot \hat{n} \, \mathrm{d}s = \int_D F \, \mathrm{d}x \, \mathrm{d}y.$$

The lemma now follows from the definition of the directional derivative and from the boundary conditions.

For future reference it is useful to observe that for harmonic functions, i.e. solutions of the Laplace equation ($F = 0$), we have

$$\int_\Gamma \partial_n u \, \mathrm{d}s = 0 \tag{2.109}$$

for any closed curve Γ that is fully contained in D.

Notice that we supplied just a single boundary condition for each one of the three problems we presented (Dirichlet, Neumann, third kind). Although we are dealing with second-order equations, the boundary conditions are quite different from the conditions we supplied in the hyperbolic case. There we provided two conditions (one on the solution and one on its derivative with respect to t) for each point on the line $t = 0$. The following example (due to Hadamard) demonstrates the difference between elliptic and hyperbolic equations on the upper half-plane. Consider Laplace's equation in the domain $-\infty < x < \infty, y > 0$, under the Cauchy conditions

$$u^n(x,0) = 0, \quad u^n_y(x,0) = \frac{\sin nx}{n} \quad -\infty < x < \infty, \tag{2.110}$$

where n is a positive integer. It is easy to check that

$$u^n(x,y) = \frac{1}{n^2} \sin nx \sinh ny$$

is a harmonic function satisfying (2.106). Choosing n to be a very large

number, the initial conditions describe an arbitrarily small perturbation of the trivial solution $u = 0$. On the other hand, the solution is not bounded at all in the half-plane $y > 0$. In fact, for any $y > 0$, the value of $\sup_{x\in\mathbb{R}} |u^n(x, y)|$ grows exponentially fast as $n \to \infty$. Thus the Cauchy problem for the Laplace equation is not stable and hence is not well posed with respect to the initial conditions (2.106).

Before developing a general theory, let us compute some special harmonic functions defined over the entire plane (except, maybe, for certain isolated points). We define a harmonic polynomial of degree n to be a harmonic function $P_n(x, y)$ of the form

$$P_n(x, y) = \sum_{0\le i+j\le n} a_{i,j} x^i y^j$$

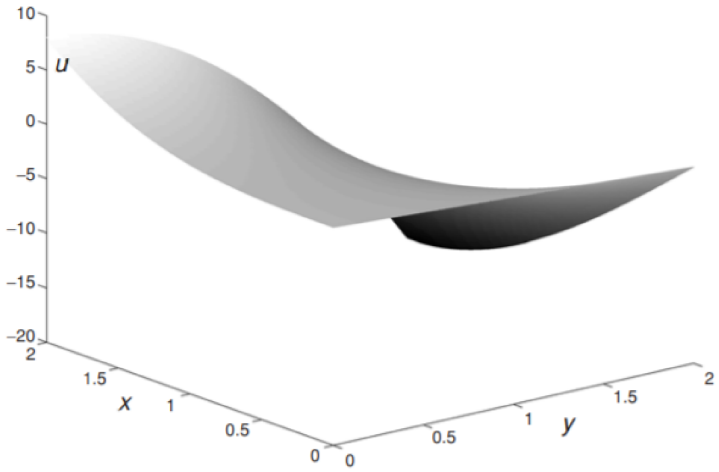

Figure 2.7: The surface of the harmonic polynomial $u(x, y) = x^3 - 3xy^2 - y$. Harmonic functions often have a saddle-like shape. This is a consequence of the maximum principle that we prove in Theorem 7.5.

For example, the functions $x - y, x^2 - y^2 + 2x, x^3 - 3xy^2 - y$ are harmonic polynomials of degree 1,2 and 3 respectively. The graph of the harmonic polynomial $u(x, y) = x^3 - 3xy^2 - y$ is depicted in Figure 2.7. The subclass V_n of harmonic polynomials P_n^{H} of the form

$$P_n^{\mathrm{H}} = \sum_{i+j=n} a_{i,j} x^i y^j$$

is called the set of homogeneous harmonic polynomials of order n. We can that for each $n > 0$ the dimension of the space V_n is exactly 2 (this result holds only in $\mathbb{R}^2$).

The most important solution of the Laplace equation over the plane is the solution that is symmetric about the origin (the radial solution). To find this solution it is convenient to use polar coordinates. We denote the polar variables by (r, θ), and the harmonic function by $w(r, \theta) = u(x(r, \theta), y(r, \theta))$. We can that the Laplace equation takes the following form in polar coordinates:

$$\Delta w = w_{rr} + \frac{1}{r} w_r + \frac{1}{r^2} w_{\theta\theta} = 0. \tag{2.111}$$

Therefore the radial symmetric solution $w(r)$ satisfies

$$w'' + \frac{1}{r} w' = 0, \tag{2.112}$$

which is Euler (equidimensional) second-order ODE . One solution is the constant function (a harmonic polynomial of degree 0 , in fact), and the other solution is given by

$$w(r) = -\frac{1}{2\pi} \ln r. \tag{2.113}$$

The solution $w(r)$ in (2.109) is called the fundamental solution of the Laplace equation. We shall use this solution extensively in Chapter 8 , where the title 'fundamental' will be justified. We shall also see there the reason for including the multiplicative constant $-1/2\pi$. Notice that the fundamental solution is not defined at the origin. The fundamental solution describes the electric potential due to a point-like electric charge at the origin.

It is interesting to note that the Laplace equation is symmetric with respect to coordinate shift: i.e. if $u(x, y)$ is a harmonic function, then so is $u(x - a, y - b)$ for any constants a and b. There are other symmetries as well; for example, the equation is symmetric with respect to rotations of the coordinate system, i.e. if $w(r, \theta)$ is harmonic, then $w(r, \theta + \gamma)$ is harmonic too for every constant γ. Another important symmetry concerns dilation of the coordinate system: if $u(x, y)$ is harmonic, then $u(x/\delta, y/\delta)$ is also harmonic for every positive constant δ.

2.2.3 The maximum principle

One of the central tools in the theory of (second-order) elliptic PDEs is the maximum principle. We first present a 'weak' form of this principle.

Theorem 2.2.2. *(The weak maximum principle) Let D be a bounded domain, and let $u(x, y) \in C^2(D) \cap C(\bar{D})$ be a harmonic function in D. Then the maximum of u in $\bar{D}$ is achieved on the boundary ∂D.*

Proof. Consider a function $v(x, y) \in C^2(D) \cap C(\bar{D})$ satisfying $\Delta v > 0$ in D. We argue that v cannot have a local maximum point in D. To see why, recall from calculus that if $(x_0, y_0) \in D$ is a local maximum point of v, then $\Delta v \leq 0$, in contradiction to our assumption.

Since u is harmonic, the function $v(x,y) = u(x,y) + \varepsilon\left(x^2+y^2\right)$ satisfies $\Delta v > 0$ for any $\varepsilon > 0$. Set $M = \max_{\partial D} u$, and $L = \max_{\partial D}\left(x^2+y^2\right)$. From our argument about v it follows that $v \leq M + \varepsilon L$ in D. Since $u = v - \varepsilon\left(x^2+y^2\right)$, it now follows that $u \leq M + \varepsilon L$ in D. Because ε can be made arbitrarily small, we obtain $u \leq M$ in D. □

Remark: If u is harmonic in D, then $-u$ is harmonic there too. But for any set A and for any function u we have

$$\min_A u = -\max_A(-u)$$

Therefore the minimum of a harmonic function u is also obtained on the boundary ∂D.

The theorem we have just proved still does not exclude the possibility that the maximum (or minimum) of u is also attained at an internal point. We shall now prove a stronger result that asserts that if u is not constant, then the maximum (and minimum) cannot, in fact, be obtained at any interior point. For this purpose we need first to establish one of the marvelous properties of harmonic functions.

Theorem 2.2.3. *(The mean value principle) Let D be a planar domain, let u be a harmonic function there and let (x_0, y_0) be a point in D. Assume that B_R is a disk of radius R centered at (x_0, y_0), fully contained in D. For any $r > 0$ set $C_r = \partial B_r$. Then the value of u at (x_0, y_0) is the average of the values of u on the circle C_R :*

$$\begin{aligned} u(x_0,y_0) &= \frac{1}{2\pi R}\oint_{C_R} u(x(s),y(s))\mathrm{d}s \\ &= \frac{1}{2\pi}\int_0^{2\pi} u\left(x_0 + R\cos\theta, y_0 + R\sin\theta\right)\mathrm{d}\theta \end{aligned} \qquad (2.114)$$

Proof. Let $0 < r \leq R$. We write $v(r,\theta) = u\left(x_0 + r\cos\theta, y_0 + r\sin\theta\right)$. We also define the integral of v with respect to θ :

$$V(r) = \frac{1}{2\pi r}\oint_{C_r} v \, \mathrm{d}s = \frac{1}{2\pi}\int_0^{2\pi} v(r,\theta)\mathrm{d}\theta.$$

Differentiating with respect to r we obtain

$$\begin{aligned} V_r(r) &= \tfrac{1}{2\pi}\textstyle\int_0^{2\pi} v_r(r,\theta)\mathrm{d}\theta \\ &= \tfrac{1}{2\pi}\textstyle\int_0^{2\pi} \frac{\partial}{\partial r} u\left(x_0 + r\cos\theta, y_0 + r\sin\theta\right)\mathrm{d}\theta \\ &= \tfrac{1}{2\pi r}\textstyle\oint_{C_r} \partial_n u \, \mathrm{d}s \\ &= 0, \end{aligned} \tag{2.115}$$

where in the last equality we used (7.9). Hence $V(r)$ does not depend on r, and thus

$$u\left(x_0, y_0\right) = V(0) = \lim_{\rho\to 0} V(\rho) = V(r) = \frac{1}{2\pi r}\oint_{C_r} u(x(s), y(s))\mathrm{d}s$$

for all $0 < r \leq R$. □

Remark It is interesting to note that the reverse statement is also true, i.e. a continuous function that satisfies the mean value property in some domain D is harmonic in D.

We prove next a slightly weaker result.

Theorem 2.2.4. *Let u be a function in $C^2(D)$ satisfying the mean value property at every point in D. Then u is harmonic in D.*

Proof. Assume by contradiction that there is a point (x_0, y_0) in D where $\Delta u\left(x_0, y_0\right) \neq 0$. Without loss of generality assume $\Delta u\left(x_0, y_0\right) > 0$.

Since $\Delta u(x, y)$ is a continuous function, then for a sufficiently small $R > 0$ there exists in D a disk B_R of radius R, centered at (x_0, y_0), such that $\Delta u > 0$ at each point in B_R. Denote the boundary of this disk by C_R. It follows that

$$\begin{aligned} 0 &< \tfrac{1}{2\pi} \int_{B_R} \Delta u \, \mathrm{d}x \, \mathrm{d}y \\ &= \tfrac{1}{2\pi} \oint_{C_R} \partial_n u \, \mathrm{d}s \\ &= \tfrac{R}{2\pi} \int_0^{2\pi} \tfrac{\partial}{\partial R} u \left(x_0 + R\cos\theta, y_0 + R\sin\theta\right) \mathrm{d}\theta \\ &= \tfrac{R}{2\pi} \tfrac{\partial}{\partial R} \int_0^{2\pi} u \left(x_0 + R\cos\theta, y_0 + R\sin\theta\right) \mathrm{d}\theta \\ &= R \tfrac{\partial}{\partial R} \left[u \left(x_0, y_0\right)\right] \\ &= 0, \end{aligned} \tag{2.116}$$

where in the fourth equality in (2.112) we used the assumption that u satisfies the mean value property. □

As a corollary of the mean value theorem, we shall prove another maximum principle for harmonic functions.

Theorem 2.2.5. ***(The strong maximum principle)*** *Let u be a harmonic function in a domain D (here we also allow for unbounded D). If u attains it maximum (minimum) at an interior point of D, then u is constant.*

Proof. Assume by contradiction that u obtains its maximum at some interior point q_0. Let $q \neq q_0$ be an arbitrary point in D. Denote by l a smooth orbit in D connecting q_0 and q (see Figure 2.8). In addition, denote by d_l the distance between l and ∂D.

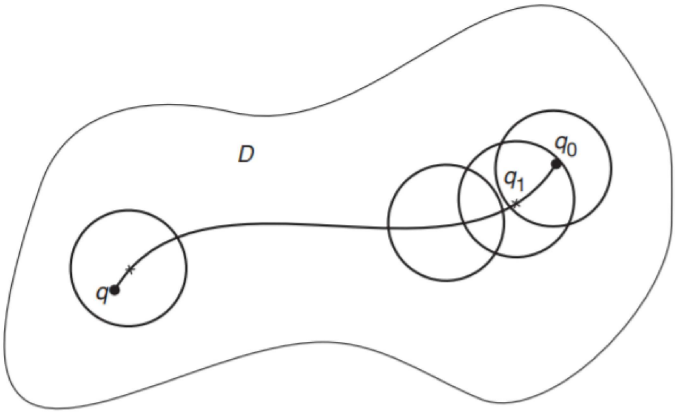

Figure 2.8: A construction for the proof of the strong maximum principle

Consider a disk B_0 of radius $d_l/2$ around q_0. From the definition of d_l and from the mean value theorem, we infer that u is constant in B_0 (since the average of a set cannot be greater than all the objects of the set). Select now a point q_1 in $l \cap B_0$, and denote by B_1 the disk of radius $d_l/2$ centered at q_1. From our construction it follows that u also reaches its maximal value at q_1. Thus we obtain that u is constant also in B_1.

We continue in this way until we reach a disk that includes the point q. We conclude $u(q) = u(q_0)$, and since q is arbitrary, it follows that u is constant in D. Notice that we may choose the points $q_0, q_1, \ldots$, such that the process involves a finite number of disks $B_0, B_1, \ldots, B_{n_l}$ because the length of l is finite, and because all the disks have the same radius. □

Remark: The strong maximum theorem indeed guarantees that nonconstant harmonic functions cannot obtain their maximum or minimum in D. Notice that in unbounded domains the maximum (mini-

mum) of u is not necessarily obtained in $\bar{D}$. For example, the function $\log\left(x^2+y^2\right)$ is harmonic and positive outside the unit disk, and it vanishes on the domain's boundary. We also point out that the first proof of the maximum principle can be readily generalized to a large class of elliptic problems, while the mean value principle holds only for harmonic functions.

2.2.4 Applications of the maximum principle

We shall illustrate the importance of the maximum principle by using it to prove the uniqueness and stability of the solution to the Dirichlet problem.

Theorem 2.2.6. *Consider the Dirichlet problem in a bounded domain:*

$$\begin{aligned} \Delta u &= f(x,y) \quad (x,y) \in D, \\ u(x,y) &= g(x,y) \quad (x,y) \in \partial D. \end{aligned}$$

The problem has at most one solution in $C^2(D) \cap C(\bar{D})$.

Proof. Assume by contradiction that there exist two solutions u_1 and u_2. Denote their difference by $v = u_1 - u_2$. The problem's linearity implies that v is harmonic in D, and that it vanishes on ∂D. The weak maximum principle implies, then, $0 \leq v \leq 0$. Thus $v \equiv 0$. □

We note that the boundedness of D is essential. Consider, for instance, the following Dirichlet problem:

$$\Delta u = 0 \qquad x^2 + y^2 > 4, \tag{2.117}$$

$$u(x,y) = 1 \qquad x^2 + y^2 = 4. \tag{2.118}$$

It is easy to verify that the functions $u_1 \equiv 1$ and $u_2(x,y) = \left(\ln\sqrt{x^2+y^2}\right)/\ln 2$ both solve the problem.

Theorem 2.2.7. *Let D be a bounded domain, and let u_1 and u_2 be functions in $C^2(D) \cap C(\bar{D})$ that are solutions of the Poisson equation $\Delta u = f$ with the Dirichlet conditions g_1 and g_2, respectively. Set $M_g = \max_{\partial D} |g_1(x,y) - g_2(x,y)|$. Then*

$$\max_D |u_1(x,y) - u_2(x,y)| \leq M_g.$$

Proof. Define $v = u_1 - u_2$. The construction implies that v is harmonic in D satisfying, $v = g_1 - g_2$ on ∂D. Therefore the maximum (and minimum) principle implies

$$\min_{\partial D}(g_1 - g_2) \leq v(x,y) \leq \max_{\partial D}(g_1 - g_2) \quad \forall (x,y) \in D,$$

and the theorem follows. □

2.2.5 Green's identities

We now develop another important tool for the analysis of elliptic problems Green's identities. We shall use this tool to provide an alternative uniqueness proof for the Dirichlet problem, and, in addition, we shall prove the uniqueness of solutions to the Neumann problem and to problems of the third kind. The Green's identities method is similar to the energy method we used in secton 2.1 , and to Green's formula.

Our starting point is Gauss' (the divergence) theorem:

$$\int_D \vec{\nabla} \cdot \vec{\psi}(x,y)\mathrm{d}x\ \mathrm{d}y = \int_{\partial D} \vec{\psi}(x(s),y(s)) \cdot \hat{n}\ \mathrm{d}s.$$

This theorem holds for any vector field $\vec{\psi} \in C^1(D) \cap C(\bar{D})$ and any bounded piecewise smooth domain D. Let u and v be two arbitrary functions in $C^2(D) \cap C^1(\bar{D})$. We consider several options for ψ in Gauss' theorem. Selecting

$$\vec{\psi} = \vec{\nabla} u,$$

we obtain (as we verified earlier)

$$\int_D \Delta u \, \mathrm{d}x \, \mathrm{d}y = \int_{\partial D} \partial_n u \, \mathrm{d}s. \tag{2.119}$$

The selection $\vec{\psi} = v\vec{\nabla}u - u\vec{\nabla}v$ leads to

$$\int_D (v\Delta u - u\Delta v) \mathrm{d}x \, \mathrm{d}y = \int_{\partial D} (v\partial_n u - u\partial_n v) \, \mathrm{d}s. \tag{2.120}$$

A third Green's identity

$$\int_D \vec{\nabla} u \cdot \vec{\nabla} v \, \mathrm{d}x \, \mathrm{d}y = \int_{\partial D} v\partial_n u \, \mathrm{d}s - \int_D v\Delta u \, \mathrm{d}x \, \mathrm{d}y, \tag{2.121}$$

is given as an exercise.

We applied the first Green's identity (2.115) to prove the mean value principle. We next apply the third Green's identity (2.117) to establish the general uniqueness theorem for Poisson's equation.

Theorem 2.2.8. *Let D be a smooth domain.*

(a) The Dirichlet problem has at most one solution.

(b) If $\alpha \geq 0$, then the problem of the third kind has at most one solution.

(c) If u solves the Neumann problem, then any other solution is of the form $v = u + c$, where $c \in \mathbb{R}$.

Proof. We start with part (b) (part (a) is a special case of part (b)). Suppose u_1 and u_2 are two solutions of the problem of the third kind. Set $v = u_1 - u_2$. It is easy to see that v is a harmonic function in D, satisfying on ∂D the boundary condition

$$v + \alpha \partial_n v = 0.$$

Substituting $v = u$ in the third Green's identity (2.117), we obtain

$$\int_D |\vec{\nabla} v|^2 \, \mathrm{d}x \, \mathrm{d}y = - \int_{\partial D} \alpha \left(\partial_n v\right)^2 \, \mathrm{d}s \tag{2.122}$$

Since the left hand side of (2.118) is nonnegative, and the right hand side is nonpositive, it follows that both sides must vanish. Hence $\nabla v = 0$ in D and $\alpha \partial_n v = -v = 0$ on ∂D. Therefore v is constant in D and it vanishes on ∂D. Thus $v \equiv 0$, and $u_1 \equiv u_2$.

The proof of part (c) is similar. We first notice that one cannot expect uniqueness in the sense of parts (a) and (b), since if u is a solution to the Neumann problem, then $u + c$ is a solution too for any constant c. Indeed, we now obtain from the identity (2.117)

$$\int_D |\vec{\nabla} v|^2 \, \mathrm{d}x \, \mathrm{d}y = 0,$$

implying that v is constant. On the other hand, since we have no constraint on the value of v on ∂D, we cannot determine the constant. We thus obtain

$$u_1 - u_2 = \text{ constant.}$$

□

2.2.6 The maximum principle for the heat equation

The maximum principle also holds for parabolic equations. Consider the heat equation for a function $u(x, y, z, t)$ in a three-dimensional bounded domain D :

$$u_t = k\Delta u \quad (x, y, z) \in D \quad t > 0, \tag{2.123}$$

where here we write $\Delta u = u_{xx} + u_{yy} + u_{zz}$. To formulate the maximum principle we define the domain

$$Q_T = \{(x, y, z, t) \mid (x, y, z) \in D, 0 < t \leq T\}.$$

Notice that the time interval $(0, T)$ is arbitrary. It is convenient at this stage to define the parabolic boundary of Q_T :

$$\partial_P Q_T = \{D \times \{0\}\} \cup \{\partial D \times [0, T]\},$$

that is the boundary of Q_T, save for the top cover $D \times \{T\}$. We also denote by C_H the class of functions that are twice differentiable in Q_T with respect to (x, y, z), once differentiable with respect to t, and continuous in $\bar{Q}_T$. We can now state the (weak) maximum principle for the heat equation.

Theorem 2.2.9. *Let $u \in C_H$ be a solution to the heat equation (2.119) in Q_T. Then u achieves its maximum (minimum) on $\partial_P Q_T$.*

Proof. We prove the statement with respect to the maximum of u. The proof with respect to the minimum of u follows at once, since if u satisfies the heat equation, so does $-u$. It is convenient to start with the following

proposition. □

Proposition 2.2.10. *Let v be a function in C_H satisfying $v_t - k\Delta v < 0$ in Q_T. Then v has no local maximum in Q_T. Moreover, v achieves its maximum in $\partial_P Q_T$.*

Proof. If v has a local maximum at some $q \in Q_T$, then $v_t(q) = 0$, implying $\Delta v(q) > 0$, which contradicts the assumption. Since v is continuous in the closed domain $\bar{Q}_T$, its maximum is achieved somewhere on the boundary $\partial(Q_T)$. If the maximum is achieved at a point $q = (x_0, y_0, z_0, T)$ on the top cover $D \times \{T\}$, then we must have $v_t(q) \geq 0$, and thus $\Delta v(q) > 0$. Again this contradicts the assumption on v, since (x_0, y_0, z_0) is a local maximum in D.

Returning to the maximum principle, we define (for $\varepsilon > 0$)

$$v(x, y, z, t) = u(x, y, z, t) - \varepsilon t.$$

Obviously

$$\max_{\partial_P Q_T} v \leq M := \max_{\partial_P Q_T} u \tag{2.124}$$

Since u satisfies the heat equation, it follows that $v_t - \Delta v < 0$ in Q_T. Theorem 2.118 and (2.120) imply that $v \leq M$, hence for all points in Q_T we have $u \leq M + \varepsilon T$. Because ε can be made arbitrarily small, we obtain $u \leq M$. □

As a direct consequence of the maximum principle we prove the following theorem, which guarantees the uniqueness and stability of the solution to the Dirichlet problem for the heat equation.

Theorem 2.2.11. *Let u_1 and u_2 be two solutions of the heat equation*

$$u_t - k\Delta u = F(x,t) \quad (x,y,z) \in D \quad 0 < t < T, \tag{2.125}$$

with initial conditions $u_i(x,y,z,0) = f_i(x,y,z)$, and boundary conditions

$$u_i(x,y,z,t) = h_i(x,y,z,t) \quad (x,y,z) \in \partial D \quad 0 < t < T,$$

respectively. Set

$$\delta = \max_D |f_1 - f_2| + \max_{\partial D \times |t>0\}} |h_1 - h_2| .$$

Then

$$|u_1 - u_2| \leq \delta \quad (x,y,z,t) \in \bar{Q}_T.$$

Proof. Writing $w = u_1 - u_2$, the proof is the same as for the corresponding theorem for Poisson's equation. The special case $f_1 = f_2, h_1 = h_2$ implies at once the uniqueness part of the theorem. □

Corollary 2.2.11.1. *Let*

$$u(x,t) = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L} \mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \tag{2.126}$$

be the formal solution of the heat problem

$$u_t - k u_{xx} = 0 \quad 0 < x < L, t > 0, \tag{2.127}$$

$$u(0,t) = u(L,t) = 0 \quad t \geq 0, \tag{2.128}$$

$$u(x,0) = f(x) \quad 0 \leq x \leq L. \tag{2.129}$$

If the series

$$f(x) = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L}$$

converges uniformly on $[0, L]$*, then the series (2.122) converges uniformly on* $[0, L] \times [0, T]$*, and* u *is a classical solution.*

Proof. Let $\varepsilon > 0$. By the Cauchy criterion for uniform convergence there exists N_ε such that for all $N_\varepsilon \leq k \leq l$ we have

$$\left| \sum_{n=k}^{l} B_n \sin \frac{n\pi x}{L} \right| < \varepsilon \quad \forall x \in [0, L].$$

Note that

$$\sum_{n=k}^{l} B_n \sin \frac{n\pi x}{L} \mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t}$$

is a classical solution of the heat equation; hence by the maximum principle

$$\left| \sum_{n=k}^{l} B_n \sin \frac{n\pi x}{L} \mathrm{e}^{-k\left(\frac{n\pi}{L}\right)^2 t} \right| < \varepsilon \quad \forall (x, t) \in [0, L] \times [0, T].$$

Invoking again the Cauchy criterion for uniform convergence, we infer that the series of (2.122) converges uniformly on $[0, L] \times [0, T]$ to the continuous function u. In particular, u satisfies the initial and boundary conditions. Since u satisfies the heat equation on $[0, L] \times (0, T]$, it follows that u is a classical solution. □

The heat equation 'smoothes out' the initial conditions in the sense even if the initial data are not smooth, the solution is in class C^∞ for all $t > 0$. We examine now another aspect of the smoothing property of the

heat equation. For simplicity let us return to the case of a single spatial variable x, so that our domain consists of an interval: $D = (a, b)$.

Proposition 2.2.12. *Let $u(x,t)$ be the solution of*

$$u_t = ku_{xx} \quad a < x < b \quad t > 0 \tag{2.130}$$

with Neumann's initial boundary value problem

$$u(x,0) = f(x) \quad x \in (a,b), \quad u_x(a,t) = u_x(b,t) = 0 \quad 0 < t < T.$$

Assume $f \in C^1([a,b])$ satisfying $f_x(a) = f_x(b) = 0$. Set $Q_T = (a,b) \times (0,T]$ for some $T > 0$. Then u satisfies

$$\max_{Q_T} |u_x| \leq \max_{(a,b)} |f'(x)| .$$

Proof. We differentiate u with respect to x and define $w(x,t) = u_x(x,t)$. One can readily verify that w satisfies the heat equation with Dirichlet initial boundary conditions:

$$w(x,0) = f'(x) \quad x \in (a,b), \quad w(a,t) = w(b,t) = 0 \quad 0 < t < T.$$

Therefore Proposition 2.2.12 follows from the maximum principle for the heat equation. □

2.2.7 Separation of variables for elliptic problems

The method of separation of variables, introduced in section 2.1 for the heat equation and for the wave equation, can in some cases also be applied to elliptic equations. Applying the method requires certain symmetries to hold, both for the equation and for the domain under study.

We shall demonstrate how to use the method to solve the Laplace and Poisson equations in rectangles, in disks, and in circular sectors.

First, we prove the following general result.

Proposition 2.2.13. *Consider the Dirichlet problem*

$$\begin{aligned} \Delta u &= 0 && (x,y) \in D, \\ u(x,y) &= g(x,y) && (x,y) \in \partial D, \end{aligned}$$

in a bounded domain D. Let

$$u(x,y) = \sum_{n=1}^{\infty} u_n(x,y) \tag{2.131}$$

be a formal solution of the problem, such that each $u_n(x,y)$ is a harmonic function in D and continuous in $\bar{D}$. If the series (2.127) *converges uniformly on ∂D to g, then it converges uniformly on $\bar{D}$ and u is a classical solution of the problem.*

Proof. Let $\varepsilon > 0$. By the Cauchy criterion for uniform convergence, there exists N_ε such that for all $N_\varepsilon \le k \le l$ we have

$$\left| \sum_{n=k}^{l} u_n(x,y) \right| < \varepsilon$$

for all $(x,y) \in \partial D$. By the weak maximum principle

$$\left| \sum_{n=k}^{l} u_n(x,y) \right| < \varepsilon$$

for all $(x,y) \in \partial D$. Invoking again the Cauchy criterion for uniform convergence, we infer that the series of (2.127) converges uniformly on

$\bar{D}$ to the continuous function u. In particular, u satisfies the boundary condition. Since each u_n satisfies the mean value property, the uniform convergence implies that u also satisfies the mean value property, and by above remark u is harmonic in D. □

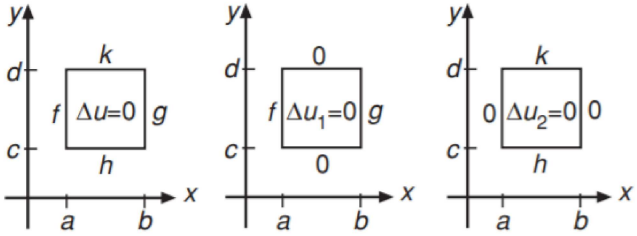

Figure 2.9: Separation of variables in rectangles

Rectangles

Let u be the solution to the Dirichlet problem in a rectangular domain D (Figure 2.9):

$$\Delta u = 0 \quad a < x < b, \quad c < y < d, \tag{2.132}$$

with the boundary conditions

$$u(a, y) = f(y), u(b, y) = g(y), u(x, c) = h(x), u(x, d) = k(x). \tag{2.133}$$

We recall that the method of separation of variables is based on constructing an appropriate eigenvalue (Sturm-Liouville) problem. This,

in turn, requires homogeneous boundary conditions. We thus split u into $u = u_1 + u_2$, where u_1 and u_2 are both harmonic in D, and where u_1 and u_2 satisfy the boundary conditions (see Figure 2.9)

$$u_1(a, y) = f(y), u_1(b, y) = g(y), u_1(x, c) = 0, u_1(x, d) = 0 \qquad (2.134)$$

and

$$u_2(a, y) = 0, u_2(b, y) = 0, u_2(x, c) = h(x), u_2(x, d) = k(x). \qquad (2.135)$$

We assume at this stage that the compatibility condition

$$f(c) = f(d) = g(c) = g(d) = h(a) = h(b) = k(a) = k(b) = 0 \qquad (2.136)$$

holds. Since $u_1 + u_2$ satisfy (2.128) and the boundary conditions (2.129), the uniqueness theorem guarantees that $u = u_1 + u_2$.

The advantage of splitting the problem into two problems is that each one of the two new problems can be solved by separation of variables. Consider, for example, the problem for u_1. We shall seek a solution in the form of a sum of separated (nonzero) functions $U(x, y) = X(x)Y(y)$. Substituting such a solution into the Laplace equation (2.128), we obtain

$$X''(x) - \lambda X(x) = 0 \qquad a < x < b, \qquad (2.137)$$

$$Y''(y) + \lambda Y(y) = 0 \qquad c < y < d. \qquad (2.138)$$

The homogeneous boundary conditions imply that

$$Y(c) = Y(d) = 0. \qquad (2.139)$$

Thus, we obtain a Sturm-Liouville problem for $Y(y)$. Solving (2.134)-(2.135), we derive a sequence of eigenvalues λ_n and eigenfunctions $Y_n(y)$. We can then substitute the sequence λ_n in (2.133) and obtain an associated sequence $X_n(x)$. The general solution u_1 is written formally as

$$u_1(x,y) = \sum_n X_n(x)Y_n(y).$$

The remaining boundary conditions for u_1 will be used to eliminate the two free parameters associated with X_n for each n, just as was done in section 2.1. Instead of writing a general formula, we find it simpler to demonstrate the method via an example.

Example 2.6 Solve the Laplace equation in the rectangle $0 < x < b, 0 < y < d$, subject to the Dirichlet boundary conditions

$$u(0,y) = f(y), u(b,y) = g(y), u(x,0) = 0, u(x,d) = 0 \tag{2.140}$$

Recalling the notation u_1 and u_2 we introduced above, this problem gives rise to a Laplace equation with zero Dirichlet conditions for u_2. Therefore the uniqueness theorem implies $u_2 \equiv 0$. For u_1 we construct a solution consisting of an infinite combination of functions of the form $w(x,y) = X(x)Y(y)$. We thus obtain for $Y(y)$ the following Sturm-Liouville problem:

$$Y''(y) + \lambda Y(y) = 0 \quad 0 < y < d, \quad Y(0) = Y(d) = 0. \tag{2.141}$$

This problem was solved in section 2.1 . The eigenvalues and eigenfunctions are

$$\lambda_n = \left(\frac{n\pi}{d}\right)^2, \quad Y_n(y) = \sin\frac{n\pi}{d}y \quad n = 1, 2, \ldots \tag{2.142}$$

The equation for the x-dependent factor is

$$X''(x) - \left(\frac{n\pi}{d}\right)^2 X(x) = 0 \quad 0 < x < b. \tag{2.143}$$

To facilitate the expansion of the boundary condition into a Fourier series, we select for (2.139) the fundamental system of solutions $\{\sinh[(n\pi/d)x], \sinh[(n\pi/d)(x- \text{ b})]\}$. Thus we write for u_1

$$u_1(x,y) = \sum_{n=1}^{\infty} \sin\frac{n\pi y}{d}\left[A_n \sinh\frac{n\pi}{d}x + B_n \sinh\frac{n\pi}{d}(x-b)\right]. \tag{2.144}$$

Substituting the expansion (2.140) into the nonhomogeneous boundary conditions of (2.136) we obtain:

$$g(y) = \sum_{n=1}^{\infty} A_n \sinh\frac{n\pi b}{d}\sin\frac{n\pi y}{d},$$

$$f(y) = \sum_{n=1}^{\infty} B_n \sinh\frac{-n\pi b}{d}\sin\frac{n\pi y}{d}.$$

To evaluate the sequences $\{A_n\}, \{B_n\}$, expand $f(y)$ and $g(y)$ into generalized Fourier series:

$$f(y) = \sum_{n=1}^{\infty} \alpha_n \sin\frac{n\pi y}{d}, \quad \alpha_n = \frac{2}{d}\int_0^d f(y)\sin\frac{n\pi y}{d}\,\mathrm{d}y,$$
$$g(y) = \sum_{n=1}^{\infty} \beta_n \sin\frac{n\pi y}{d}, \quad \beta_n = \frac{2}{d}\int_0^d g(y)\sin\frac{n\pi y}{d}\,\mathrm{d}y.$$

This implies

$$A_n = \frac{\beta_n}{\sinh\frac{n\pi b}{d}}, \quad B_n = -\frac{\alpha_n}{\sinh\frac{n\pi b}{d}}.$$

We saw in section 2.1 that the generalized Fourier series representing

the solution to the heat equations converges exponentially fast for all $t > 0$. Moreover, the series for the derivatives of all orders converges too. On the other hand, the rate of convergence for the series representing the solution to the wave equation depends on the smoothness of the initial data, and singularities in the initial data are preserved by the solution. What is the convergence rate for the formal series representing the solution of the Laplace equation? Do we obtain a classical solution?

To answer these questions, let us consider the general term in the series (2.140). We assume that the functions $f(y)$ and $g(y)$ are piecewise differentiable, and that they satisfy the homogeneous Dirichlet conditions at the end points $y = 0, d$. Then the coefficients α_n and β_n satisfy $|\alpha_n| < C_1, |\beta_n| < C_2$, where C_1 and C_2 are constants that do not depend on n (actually, one can establish a far stronger result on the decay of α_n and β_n to zero as $n \to \infty$, but we do not need this result here). Consider a specific term

$$\frac{\alpha_n}{\sinh \frac{n\pi b}{d}} \sinh \frac{n\pi}{d}(x-b) \sin \frac{n\pi y}{d}$$

in the series that represents u_1, where (x, y) is some interior point in D. This term is of the order of $O\left(\mathrm{e}^{-\frac{n\pi}{d}x}\right)$ for large values of n. The same argument implies

$$\frac{\beta_n}{\sinh \frac{n\pi b}{d}} \sinh \frac{n\pi}{d} x \sin \frac{n\pi y}{d} = O\left(\mathrm{e}^{\frac{n\pi}{d}(x-b)}\right).$$

Thus all the terms in the series (2.140) decay exponentially fast as $n \to \infty$. Similarly the series of derivatives of all orders also converges exponentially fast, since the k th derivative introduces an algebraic factor n^k into the nth term in the series, but this factor is negligible (for large n) in comparison with the exponentially decaying term. We point out, though, that the rate of convergence slows down as we approach the

domain's boundary.

Example 2.7 Solve the Laplace equation in the square $0 < x, y < \pi$ subject to the Dirichlet condition

$$u(x, 0) = 1984, u(x, \pi) = u(0, y) = u(\pi, y) = 0.$$

The problem involves homogeneous boundary conditions on the two boundaries parallel to the y axis, and a single nonhomogeneous condition on the boundary $y = 0$. Therefore we write the formal solution in the form

$$u(x, y) = \sum_{n=1}^{\infty} A_n \sin nx \sinh n(y - \pi).$$

Substituting the boundary condition $u(x, 0) = 1984$ we obtain

$$-\sum_{1}^{\infty} A_n \sin nx \sinh n\pi = 1984$$

The generalized Fourier series of the constant function $f(x) = 1984$ is $1984 = \sum_{n=1}^{\infty} \alpha_n \sin nx$, where the coefficients $\{\alpha_n\}$ are given by

$$\alpha_n = \frac{2 \times 1984}{\pi} \int_0^{\pi} \sin nx \, \mathrm{d}x = \frac{3968}{n\pi}(\cos 0 - \cos n\pi).$$

We thus obtain

$$A_n = \begin{cases} -\frac{7936}{(2k-1)\pi \sinh(2k-1)\pi} & n = 2k - 1, \\ 0 & n = 2k. \end{cases} \tag{2.145}$$

The solution to the problem is given formally by

$$u(x,y) = 7936 \sum_{n=1}^{\infty} \frac{\sin(2n-1)x \sinh(2n-1)(\pi - y)}{(2n-1)\pi \sinh(2n-1)\pi}. \tag{2.146}$$

The observant reader might have noticed that in the last example we violated the compatibility condition (2.132). This is the condition that guarantees that each of the problems we solve by separation of variables has a continuous solution. Nevertheless, we obtained a formal solution, and following our discussion above, we know that the solution converges at every interior point in the square. Furthermore, the convergence of the series, and all its derivatives at every interior point is exponentially fast. Why should we be bothered, then, by the violation of the compatibility condition?

The answer lies in the Gibbs phenomenon . If we compute the solution near the problematic points $((0,0),(1,0)$ in the last example) by summing up finitely many terms in the series (2.142), we observe high frequency oscillations. The difficulty we describe here is relevant not only to analytical solutions, but also to numerical solutions. We emphasize that even if the boundary condition to the original problem leads to a continuous solution, the process of breaking the problem into several subproblems for the purpose of separating variables might introduce discontinuities into the subproblems!

We therefore present a method for transforming a Dirichlet problem with con tinuous boundary data that does not satisfy the compatibility condition (2.132) into another Dirichlet problem (with continuous boundary data) that does satisfy (2.132). Denote the harmonic function we seek by $u(x,y)$, and the Dirichlet boundary condition on the rectangle's boundary by g. We write u as a combination: $u(x,y) =$

$v(x,y) - P_2(x,y)$, where P_2 is a second-order appropriate harmonic polynomial (that we still have to find), while v is a harmonic function. We construct the harmonic polynomial in such a way that v satisfies the compatibility condition (2.132) (i.e. v vanishes at the square's vertices). Denote the restriction of v and P_2 to the square's boundary by g_1 and g_2, respectively. We select P_2 so that the incompatibility of g at the square's vertices is included in g_2. Thus we obtain for v a compatible Dirichlet problem. To construct P_2 as just described, we write the general form of a second-order harmonic polynomial:

$$P_2(x,y) = a_1\left(x^2 - y^2\right) + a_2 xy + a_3 x + a_4 y + a_5. \tag{2.147}$$

For simplicity, and without loss of generality, consider the square $0 < x < 1, 0 < y < 1$. Requiring g_1 to vanish at all the vertices leads to four equations for the five unknown coefficients of P_2 :

$$\begin{aligned} g(0,0) + a_5 &= 0, \\ g(1,0) + a_1 + a_3 + a_5 &= 0, \\ g(0,1) - a_1 + a_4 + a_5 &= 0, \\ g(1,1) + a_2 + a_3 + a_4 + a_5 &= 0. \end{aligned} \tag{2.148}$$

We choose arbitrarily $a_1 = 0$ and obtain the solution:

$$\begin{aligned} a_1 &= 0 \\ a_2 &= -g(1,1) - g(0,0) + g(1,0) + g(0,1) \\ a_3 &= g(0,0) - g(1,0) \\ a_4 &= g(0,0) - g(0,1) \\ a_5 &= -g(0,0) \end{aligned} \tag{2.149}$$

Having thrown all the incompatibilities into the (easy to compute) harmonic polynomial, it remains to find a harmonic function v that satisfies the compatible boundary conditions $g_1 = g + g_2$.

Example 2.8: Let $u(x, y)$ be the harmonic function in the unit square satisfying the Dirichlet conditions

$$u(x, 0) = 1 + \sin \pi x, \quad u(x, 1) = 2, \quad u(0, y) = u(1, y) = 1 + y.$$

Represent u as a sum of a harmonic polynomial, and a harmonic function $v(x, y)$ that satisfies the compatibility condition (2.132).

We compute the appropriate harmonic polynomial. Solving the algebraic system (2.145) we get

$$a_1 = a_2 = a_3 = 0, \quad a_4 = -1, \quad a_5 = -1.$$

Hence the harmonic polynomial is $P_2(x, y) = -1 - y$. Define now

$$v(x, y) = u(x, y) - (1 + y).$$

Our construction implies that v is the harmonic function satisfying the Dirichlet data

$$v(x, 0) = \sin \pi x, \quad v(x, 1) = v(0, y) = v(1, y) = 0.$$

Indeed, the compatibility condition holds for v. Finally, we obtain that $v(x, y) = \sin \pi x \sinh(\pi - y)/\sinh \pi$, and therefore

$$u(x, y) = \sin \pi x \frac{\sinh(\pi - y)}{\sinh \pi} + 1 + y.$$

We end this section by solving a Neumann problem in a square.

Example 2.8: Find a harmonic function $u(x, y)$ in the square $0 < x$, $y < \pi$ satisfying the Neumann boundary conditions

$$u_y(x, \pi) = x - \pi/2, u_x(0, y) = u_x(\pi, y) = u_y(x, 0) = 0. \tag{2.150}$$

The first step in solving a Neumann problem is to verify that the necessary condition for existence holds. In the current case, the integral $\int_{\partial D} \partial_n u \, \mathrm{d}s$ is equal to $\int_0^\pi (x - \pi/2)\mathrm{d}x$, which indeed vanishes.

The nature of the boundary conditions implies separated solutions of the form $U_n(x, y) = \cos nx \cosh ny$, where $n = 0, 1, 2 \ldots$. Thus the formal solution is

$$u(x, y) = A_0 + \sum_{n=1}^{\infty} A_n \cos nx \cosh ny. \tag{2.151}$$

The function u represented by (2.147) formally satisfies the equation and all the homogeneous boundary conditions. Substituting this solution into the nonhomogeneous boundary conditions on the edge $y = \pi$ leads to

$$\sum_{n=1}^{\infty} nA_n \sinh n\pi \cos nx = x - \pi/2 \quad 0 < x < \pi.$$

We therefore expand $x - \pi/2$ into the following generalized Fourier series:

$$\begin{aligned} x - \pi/2 &= \sum_{n=1}^{\infty} \beta_n \cos nx, \\ \beta_n &= \frac{2}{\pi} \int_0^\pi (x - \pi/2) \cos nx \, \mathrm{d}x \\ &= \begin{cases} -4/\pi n^2 & n = 1, 3, 5, \ldots. \\ 0 & n = 0, 2, 4, \ldots. \end{cases} \end{aligned}$$

Thus

$$u(x,y) = A_0 - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\cos(2n-1)x\cosh(2n-1)y}{(2n-1)^3\sinh(2n-1)\pi}. \tag{2.152}$$

The graph of u is depicted in Figure 2.10. Notice that the additive constant A_0 is not determined by the problem's conditions. This was expected due to the nonuniqueness of the Neumann problem .

Remark: When we considered the Dirichlet problem earlier, we sometimes had to divide the problem into two subproblems, each of which involved homogeneous Dirichlet conditions on two opposite edges of the rectangle. A similar division is sometimes needed for the Neumann problem. Here, however, a

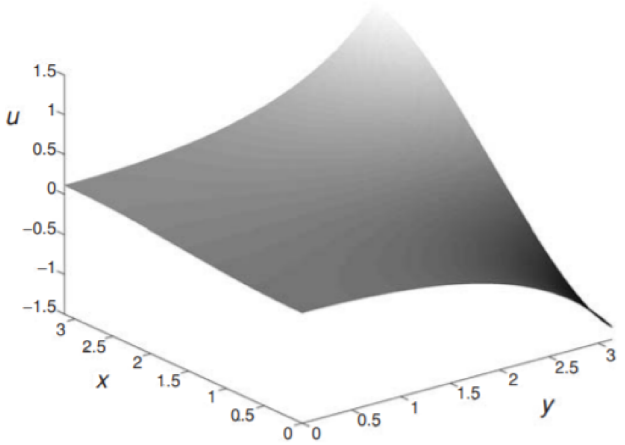

Figure 2.10: The graph of $u(x,y)$ from (2.148). Observe that in spite of the intricate form of the Fourier series, the actual shape of the surface is very smooth. We also see that u achieves its maximum and minimum on the boundary.

fundamental difficulty might arise: while the original problem presumably satisfies the necessary existence condition (otherwise the problem is not solvable at all!), it is not guaranteed that each of the subproblems will satisfy this condition! To demonstrate the difficulty and to propose a remedy for it, we look at it in some detail. Consider the Neumann problem for the Laplace equation:

$$\Delta u = 0 \qquad x \in \Omega, \tag{2.153}$$

$$\partial_n u = g \qquad x \in \partial\Omega. \tag{2.154}$$

We assume, of course, that the condition $\int_{\partial\Omega} g ds = 0$ holds. Split the boundary of Ω into two parts $\partial\Omega = \partial_1\Omega \cup \partial_2\Omega$. Define $u = u_1 + u_2$, where u_1, u_2 are both harmonic in Ω and satisfy the boundary conditions

$$\partial_n u_1 = \begin{cases} g & x \in \partial_1\Omega, \\ 0 & x \in \partial_2\Omega. \end{cases} \qquad \partial_n u_2 = \begin{cases} 0 & x \in \partial_1\Omega, \\ g & x \in \partial_2\Omega. \end{cases}$$

The difficulty is that now the existence condition may not hold separately for u_1 and u_2. We overcome this by the same method we used earlier to take care of the Gibbs phenomenon. We add to (and subtract from) the solution a harmonic polynomial. We use a harmonic polynomial $P(x, y)$ that satisfies $\int_{\partial_1\Omega} \partial_n P(x, y) \mathrm{d}s \neq 0$. Assume, for example, that the harmonic polynomial $x^2 - y^2$ satisfies this condition. We then search for harmonic functions v_1 and v_2 that satisfy the following Neumann conditions:

$$\partial_n v_1 = \{ \begin{array}{ll} g + a\partial_n \left(x^2 - y^2\right) & x \in \partial_1\Omega, \\ 0 & x \in \partial_2\Omega, \end{array}$$

$$\partial_n v_2 = \begin{cases} 0 & x \in \partial_1\Omega, \\ g + a\partial_n\left(x^2 - y^2\right) & x \in \partial_2\Omega. \end{cases}.$$

We choose the parameter a such that the solvability condition holds for v_1. Since the original problem is assumed to be solvable and also the harmonic polynomial, by its very existence, satisfies the compatibility condition, it follows that v_2 must satisfy that condition too. Finally, we observe that $u = v_1 + v_2 - a\left(x^2 - y^2\right)$.

Circular domains

Another important domain where the Laplace equation can be solved by separation of variables is a disk. Let B_a be a disk of radius a around the origin. We want to find the function $u(x, y)$ that solves the Dirichlet problem

$$\Delta u = 0 \quad (x, y) \in B_a, \tag{2.155}$$

$$u(x, y) = g(x, y) \quad (x, y) \in \partial B_a. \tag{2.156}$$

It is convenient to solve the equation in polar coordinates in order to use the symmetry of the domain. We thus denote the polar coordinates by (r, θ), and the unknown function is written as $w(r, \theta) = u(x(r, \theta), y(r, \theta))$. We mentioned in Subsection 2.2.2 that w satisfies the equation

$$\Delta w = w_{rr} + \frac{1}{r}w_r + \frac{1}{r^2}w_{\theta\theta} = 0. \tag{2.157}$$

Consequently, we have to solve (2.153) in the domain

$$B_a = \{(r, \theta) \mid 0 < r < a, 0 \leq \theta \leq 2\pi\}.$$

The PDE is subject to the boundary condition

$$w(a,\theta) = h(\theta) = g(x(a,\theta), y(a,\theta)), \tag{2.158}$$

and to the additional obvious requirement that $\lim_{r\to 0} w(r,\theta)$ exists and is finite (the origin needs special attention, since it is a singular point in polar coordinates). We seek a solution of the form $w(r,\theta) = R(r)\Theta(\theta)$. Substituting this function into (2.153), and using the usual arguments from Section 2, we obtain a pair of equations for R and Θ :

$$r^2R''(r) + rR'(r) - \lambda R(r) = 0 \quad 0 < r < a, \tag{2.159}$$

$$\Theta''(\theta) + \lambda\Theta(\theta) = 0. \tag{2.160}$$

The equation for Θ holds at the interval $(0, 2\pi)$. In order that the solution $w(r,\theta)$ be of class C^2, we need to impose two periodicity conditions:

$$\Theta(0) = \Theta(2\pi), \quad \Theta'(0) = \Theta'(2\pi). \tag{2.161}$$

Notice that (2.156) and (2.157) together also imply the periodicity of the second derivative with respect to θ.

The general solution to the Sturm-Liouville problem (2.156)-(2.157) is given by the sequence

$$\Theta_n(\theta) = A_n \cos n\theta + B_n \sin n\theta, \quad \lambda_n = n^2, 0, 1, 2, \ldots \tag{2.162}$$

Substituting the eigenvalues λ_n into (2.155) yields a second-order Euler (equidimensional) ODE for R :

$$r^2R_n'' + rR_n' - n^2R_n = 0. \tag{2.163}$$

The solutions of these equations are given (except for $n = 0$) by appropriate powers of the independent variable r :

$$R_n(r) = C_n r^n + D_n r^{-n}, \quad n = 1, 2, \ldots. \tag{2.164}$$

In the special case $n = 0$ we obtain

$$R_0(r) = C_0 + D_0 \ln r. \tag{2.165}$$

Observe that the functions $r^{-n}, n = 1, 2, \ldots$ and the function $\ln r$ are singular at the origin ($r = 0$). Since we only consider smooth solutions, we impose the condition

$$D_n = 0 \quad n = 0, 1, 2, \ldots$$

We still have to satisfy the boundary condition (2.154). For this purpose we form the superposition

$$w(r, \theta) = \frac{\alpha_0}{2} + \sum_{n=1}^{\infty} r^n \left(\alpha_n \cos n\theta + \beta_n \sin n\theta\right) \tag{2.166}$$

Formally differentiating this series term-by-term, we verify that (2.162) is indeed harmonic. Imposing the boundary condition (2.154), and using the classical Fourier formula , we obtain

$$\begin{aligned} \alpha_0 &= \frac{1}{\pi} \int_0^{2\pi} h(\varphi) \mathrm{d}\varphi, \\ \alpha_n &= \frac{1}{\pi a^n} \int_0^{2\pi} h(\varphi) \cos n\varphi \mathrm{d}\varphi, \\ \beta_n &= \frac{1}{\pi a^n} \int_0^{2\pi} h(\varphi) \sin n\varphi \mathrm{d}\varphi \end{aligned} \tag{2.167}$$

$n \geq 1$.

Example 2.9 Solve the Laplace equation in the unit disk subject to the boundary conditions $w(r, \theta) = y^2$ on $r = 1$.

Observe that on the boundary $y^2 = \sin^2 \theta$. All we have to do is to compute the classical Fourier expansion of the function $\sin^2 \theta$. This expansion is readily performed in light of the identity $\sin^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$. Thus the Fourier series is finite, and the required harmonic function is $w(r, \theta) = \frac{1}{2}\left(1 - r^2 \cos 2\theta\right)$, or, upon returning to Cartesian coordinates, $u(x, y) = \frac{1}{2}\left(1 - x^2 + y^2\right)$. Before proceeding to other examples, it is worthwhile to examine the convergence properties of the formal Fourier series we constructed in (2.162). We write $M = (1/\pi) \int_0^{2\pi} |h(\theta)| d\theta$. The Fourier formulas (2.163) imply the inequalities

$$|\alpha_n|, |\beta_n| \leq M a^{-n}.$$

Hence the nth term in the Fourier series (2.162) is bounded by $2M(r/a)^n$; thus the series converges for all $r < a$, and even converges uniformly in any disk of radius $\tilde{a} < a$. In fact, we can use the same argument to show also that the series of derivatives of any order converges uniformly in any disk of radius $\tilde{a} < a$ to the appropriate derivative of the solution. Moreover, if $h(\theta)$ is a periodic piecewise differentiable and continuous function, its Fourier expansion converges uniformly. Hence w is a classical solution. Observe, however, that the rate of convergence deteriorates when we approach the boundary.

Example 2.10 Find the harmonic function $w(r, \theta)$ in the sector

$$D_\gamma = \{(r, \theta) \mid 0 < r < a, 0 < \theta < \gamma\}$$

that satisfies on the sector's boundary the Dirichlet condition

$$w(a,\theta) = g(\theta) \quad 0 \le \theta \le \gamma, \quad w(r,0) = w(r,\gamma) = 0 \quad 0 \le r \le a. \tag{2.168}$$

The process of obtaining separated solutions and an appropriate eigenvalue problem is similar to the previous case of the Laplace equation in the entire disk. Namely, we again seek solutions of the form $R(r)\Theta(\theta)$, where the Sturm-Liouville equation is again (2.156) for Θ, and the equation for the radial component is again (2.155). The difference is in the boundary condition for the Θ equation. Unlike the periodic boundary conditions that we encountered for the problem in the full disk, we now have Dirichlet boundary conditions $\Theta(0) = \Theta(\gamma) = 0$. Therefore the sequences of eigenfunctions and eigenvalues are now given by

$$\Theta_n(\theta) = A_n \sin \frac{n\pi}{\gamma}\theta, \quad \lambda_n = \left(\frac{n\pi}{\gamma}\right)^2 \quad n = 1, 2, \ldots.$$

Substituting the eigenvalues λ_n into (2.155), and keeping only the solutions that are bounded in the origin, we obtain

$$w_n(r,\theta) = \sin \frac{n\pi\theta}{\gamma} r^{n\pi/\gamma}.$$

Hence, the formal solution is given by the series

$$w(r,\theta) = \sum_{n=1}^{\infty} \alpha_n \sin \frac{n\pi\theta}{\gamma} r^{n\pi/\gamma}. \tag{2.169}$$

On $r = a, 0 < \theta < \gamma$ we have

$$g(\theta) = \sum_{n=1}^{\infty} \alpha_n a^{n\pi/\gamma} \sin \frac{n\pi\theta}{\gamma},$$

therefore,

$$\alpha_n = \frac{2a^{-n\pi/\gamma}}{\gamma} \int_0^\gamma g(\varphi) \sin \frac{n\pi\varphi}{\gamma} \mathrm{d}\varphi.$$

Remark: Consider the special case in which $\gamma = 2\pi$, and write explicitly the solution (2.165):

$$w(r,\theta) = \sum_{n=1}^{\infty} \alpha_n \sin \frac{n\theta}{2} r^{n/2}. \tag{2.170}$$

Observe that even though the sector now consists of the entire disk, the solution (2.166) is completely different from the solution we found earlier for the Dirichlet problem in the disk. The reason for the difference is that the boundary condition (2.164) on the sector's boundary is fundamentally different from the periodic boundary condition. The condition (2.164) singles out a specific curve in the disk, and, thus, breaks the disk's symmetry.

Remark: We observe that, in general, some derivatives (in fact, most of them) of the solution (2.165) are singular at the origin (the sector's vertex). We cannot require the solution to be as smooth there as we wish. This singularity has important physical significance. The Laplace equation in a sector is used to model cracks in the theory of elasticity. The singularity we observe indicates a concentration of large stresses at the vertex.

We end this section by demonstrating the separation of variables method for the Poisson equation in a disk with Dirichlet boundary conditions. The problem is thus to find a function $w(r,\theta)$, satisfying

$$\Delta w = F(r,\theta) \quad 0 < r < a, \quad 0 \le \theta \le 2\pi,$$

together with the boundary condition $w(a,\theta) = g(\theta)$. In light of the

general technique to solve nonhomogeneous equations, we seek a solution in the form

$$w(r,\theta) = \frac{f_0(r)}{2} + \sum_{n=1}^{\infty} \left[f_n(r)\cos n\theta + g_n(r)\sin n\theta \right].$$

Similarly we expand F into a Fourier series

$$F(r,\theta) = \frac{\delta_0(r)}{2} + \sum_{n=1}^{\infty} \left[\delta_n(r)\cos n\theta + \epsilon_n(r)\sin n\theta \right].$$

Substituting these two Fourier series into the Poisson equation, and comparing the associated coefficients, we find

$$f_n'' + \frac{1}{r} f_n' - \frac{n^2}{r^2} f_n = \delta_n(r) \quad n = 0, 1, \ldots, \tag{2.171}$$

$$g_n'' + \frac{1}{r} g_n' - \frac{n^2}{r^2} g_n = \epsilon_n(r) \quad n = 1, 2, \ldots \tag{2.172}$$

The general solutions of these equations can be written as

$$f_n(r) = A_n r^n + \tilde{f}_n(r), \quad g_n(r) = B_n r^n + \tilde{g}_n(r),$$

where $\tilde{f}_n$ and $\tilde{g}_n$ are particular solutions of the appropriate nonhomogeneous equations. In fact, it is shown in Exercise 7.18 that the solutions of (2.167) – (2.168), that satisfy the homogeneous boundary conditions $\tilde{f}_n(a) = \tilde{g}_n(a) = 0$, and that are bounded at the origin, can be written as

$$\tilde{f}_n(r) = \int_0^r K_1^{(n)}(r,a,\rho)\delta_n(\rho)\rho \mathrm{d}\rho + \int_r^a K_2^{(n)}(r,a,\rho)\delta_n(\rho)\rho \mathrm{d}\rho, \tag{2.173}$$

$$\tilde{g}_n(r) = \int_0^r K_1^{(n)}(r,a,\rho)\epsilon_n(\rho)\rho \mathrm{d}\rho + \int_r^a K_2^{(n)}(r,a,\rho)\epsilon_n(\rho)\rho \mathrm{d}\rho, \quad (2.174)$$

where

$$K_1^{(0)} = \log\frac{r}{a}, \quad K_2^{(0)} = \log\frac{\rho}{a}, \quad (2.175)$$

$$\begin{aligned} K_1^{(n)} &= \frac{1}{2n}\left[\left(\frac{r}{a}\right)^n - \left(\frac{a}{r}\right)^n\right]\left(\frac{\rho}{a}\right)^n, \\ K_2^{(n)} &= \frac{1}{2n}\left[\left(\frac{\rho}{a}\right)^n - \left(\frac{a}{\rho}\right)^n\right]\left(\frac{r}{a}\right)^n \quad n \geq 1. \end{aligned} \quad (2.176)$$

We thus constructed a solution of the form

$$\begin{aligned} &w(r,\theta) \\ &= \frac{A_0 + \tilde{f}_0(r)}{2} + \sum_{n=1}^{\infty}\left\{\left[A_n r^n + \tilde{f}_n(r)\right]\right. \\ &\left.\cos n\theta + [B_n r^n + \tilde{g}_n(r)]\sin n\theta\right\}. \end{aligned} \quad (2.177)$$

To find the coefficients $\{A_n, B_n\}$ we substitute this solution into the boundary conditions

$$\begin{aligned} &w(a,\theta) \\ &= \frac{A_0 + \tilde{f}_0(a)}{2} + \sum_{n=1}^{\infty}\left\{\left[A_n a^n + \tilde{f}_n(a)\right]\right. \\ &\left.\cos n\theta + [B_n a^n + \tilde{g}_n(a)]\sin n\theta\right\} \\ &= \frac{\alpha_0}{2} + \sum_{n=1}^{\infty}(\alpha_n \cos n\theta + \beta_n \sin n\theta) \\ &= g(\theta). \end{aligned} \quad (2.178)$$

The required coefficients can be written as

$$A_n = \frac{\alpha_n - 2\tilde{f}_n(a)}{a^n} = \frac{\alpha_n}{a^n} \quad n = 0, 1, \ldots,$$

$$B_n = \frac{\beta_n - \tilde{g}_n(a)}{a^n} = \frac{\beta_n}{a^n} \quad n = 1, 2, \ldots$$

Example 2.11 Solve the Poisson equation

$$\Delta w = 8r\cos\theta \quad 0 \le r < 1, 0 \le \theta \le 2\pi,$$

subject to the boundary conditions $w(1, \theta) = \cos^2\theta$. One can verify that

$$\tilde{f}_n(r) = \begin{cases} r^3 & n = 1, \\ 0 & n \neq 1, \end{cases} \tag{2.179}$$

and $\tilde{g}_n(r) = 0$ for every n, are particular solutions to the nonhomogeneous equations (2.167)-(2.168). Therefore the general solution can be written as

$$\begin{aligned} w(r,\theta) &= \frac{A_0}{2} + \left(A_1 r + r^3\right) \\ &\quad \cos\theta + B_1 r\sin\theta + \sum_{n=2}^{\infty} \left(A_n r^n \cos n\theta + B_n r^n \sin n\theta\right). \end{aligned}$$

We use the identity $\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$ to obtain the expansion of the boundary condition into a Fourier series. Therefore,

$$A_0 = 1, A_1 = -1, A_2 = \frac{1}{2}, A_n = 0 \quad \forall n \neq 0, 1, 2,$$
$$B_n = 0 \quad \forall n = 1, 2, 3 \ldots,$$

and the solution is

$$w(r,\theta) = \frac{1}{2} + \left(r^3 - 1\right)\cos\theta + \frac{r^2}{2}\cos 2\theta.$$

2.2.8 Poisson's formula

One of the important tools in the theory of PDEs is the integral representation of solutions. An integral representation is a formula for the solution of a problem in terms of an integral depending on a kernel function. We need to compute the kernel function just once for a given equation, a given domain, and a given type of boundary condition. We demonstrate now an integral representation for the Laplace equation in a disk of radius a with Dirichlet boundary conditions (2.151)-(2.152). We start by rewriting the solution as a Fourier series (see (2.162)), using (2.163):

$$w(r,\theta) = \frac{1}{2\pi}\int_0^{2\pi} h(\varphi)\mathrm{d}\varphi + \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n \int_0^{2\pi} h(\varphi)(\cos n\varphi\cos n\theta + \sin n\varphi \sin n\theta)\mathrm{d}\varphi. \tag{2.180}$$

Consider $r < \tilde{a} < a$. Since the series converges uniformly there, we can interchange the order of summation and integration, and obtain

$$w(r,\theta) = \frac{1}{\pi}\int_0^{2\pi} h(\varphi)\left[\frac{1}{2} + \sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n \cos n(\theta-\varphi)\right]\mathrm{d}\varphi. \tag{2.181}$$

The summation of the infinite series

$$\frac{1}{2}+\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n\cos n(\theta-\varphi)$$

requires a little side calculation. Define for this purpose $z=\rho e^{i\alpha}$ and evaluate (for $\rho<1$) the geometric sum

$$\frac{1}{2}+\sum_{1}^{\infty}z^n=\frac{1}{2}+\frac{z}{1-z}=\frac{1-\rho^2+2i\rho\sin\alpha}{2\left(1-2\rho\cos\alpha+\rho^2\right)}.$$

Since $z^n=\rho^n(\cos n\alpha+i\sin n\alpha)$, we conclude upon separating the real and imaginary parts that

$$\frac{1}{2}+\sum_{1}^{\infty}\rho^n\cos n\alpha=\frac{1-\rho^2}{2\left(1-2\rho\cos\alpha+\rho^2\right)} \tag{2.182}$$

Returning to (2.177) using $\rho=r/a,\alpha=\theta-\varphi$, we obtain the Poisson formula

$$w(r,\theta)=\frac{1}{2\pi}\int_0^{2\pi}K(r,\theta;a,\varphi)h(\varphi)d\varphi, \tag{2.183}$$

where the kernel K, given by

$$K(r,\theta;a,\varphi)=\frac{a^2-r^2}{a^2-2ar\cos(\theta-\varphi)+r^2}, \tag{2.184}$$

is called Poisson's kernel. This is a very useful formula. The kernel describes a universal solution for the Laplace equation in a disk. All we have to do (at least in theory), is to substitute the boundary condition into (2.179) and carry out the integration. Moreover, the formula is valid for any integrable function h. It turns out that one can derive similar representations not just in disks, but also in arbitrary smooth domains.

As another example of an integral representation for harmonic functions, we derive the Poisson formula for the Neumann problem in a disk. Let $w(r,\theta)$ be a harmonic function in the disk $r < a$, satisfying on the disk's boundary the Neumann condition $\partial w(a,\theta)/\partial r = g(\theta)$. We assume, of course, that the solvability condition $\int_0^{2\pi} g(\theta)\mathrm{d}\theta = 0$ holds. Recall that the general form of a harmonic function in the disk is

$$w(r,\theta) = \frac{\alpha_0}{2} + \sum_{n=1}^{\infty} r^n \left(\alpha_n \cos n\theta + \beta_n \sin n\theta\right).$$

Note that the coefficient α_0 is arbitrary, and cannot be retrieved from the boundary conditions (cf. the uniqueness theorem for the Neumann problem). To find the coefficients $\{\alpha_n, \beta_n\}$, substitute the solution into the boundary conditions and obtain

$$\begin{aligned} \alpha_n &= \frac{1}{n\pi a^{n-1}} \int_0^{2\pi} g(\varphi)\cos n\varphi \mathrm{d}\varphi, \\ \beta_n &= \frac{1}{n\pi a^{n-1}} \int_0^{2\pi} g(\varphi)\sin n\varphi \mathrm{d}\varphi \quad n = 1, 2, \ldots. \end{aligned} \tag{2.185}$$

Hence

$$w(r,\theta) = \frac{\alpha_0}{2} + \frac{a}{\pi} \int_0^{2\pi} K_N(r,\theta;a,\varphi) g(\varphi)\mathrm{d}\varphi, \tag{2.186}$$

where

$$K(r,\theta;a,\varphi) = \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{r}{a}\right)^n \cos n(\theta - \varphi). \tag{2.187}$$

Because of the $1/n$ factor we cannot use the summation formula (2.178) directly. Instead we perform another quick side calculation. Notice that

by a process like the one leading to (2.178) one can derive

$$\sum_{n=1}^{\infty} \rho^{n-1} \cos n\alpha = \frac{\cos\alpha - \rho}{1 - 2\rho\cos\alpha + \rho^2}.$$

Therefore after an integration with respect to ρ we obtain the Poisson kernel for the Neumann problem in a disk:

$$\begin{aligned} K_N(r,\theta;a,\varphi) &= \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{r}{a}\right)^n \cos n(\theta-\varphi) \\ &= -\frac{1}{2} \ln\left[1 - 2\frac{r}{a}\cos(\theta-\varphi) + \left(\frac{r}{a}\right)^2\right] \end{aligned} \tag{2.188}$$

Remark: It is interesting to note that both Poisson's kernels that we have computed have a special dependency on the angular variable θ. In both cases we have

$$K(r,\theta;a,\varphi) = \tilde{K}(r,a,\theta-\varphi),$$

namely, the dependency is only through the difference between θ and φ. This property is a consequence of the symmetry of the Laplace equation and the circular domain with respect to rotations. **Remark:** Poisson's formula provides, as a by-product, another proof for the mean value principle. In fact, the formula is valid with respect to any circle around any point in a given domain (provided that the circle is fully contained in the domain). Indeed, if we substitute $r = 0$ into (2.179), we obtain at once the mean value principle.

When we solved the Laplace equation in a rectangle or in a disk, we saw that the solution is in $C^\infty(D)$. That is, the solution is differentiable

infinitely many times at any interior point. Let us prove this property for any domain.

Theorem 2.2.14. ***(Smoothness of harmonic functions)*** *Let $u(x, y)$ be a harmonic function in D. Then $u \in C^\infty(D)$.*

Proof. Denote by p an interior point in D, and construct a coordinate system centered at p. Let B_a be a disk of radius a centered at p, fully contained in D. Write Poisson's formula for an arbitrary point (x, y) in B_a. We can differentiate under the integral sign arbitrarily many times with respect to r or with respect to θ, and thus establish the theorem. □

Exercises

Problem 2.2.1. *Prove Green's identity(2.2.13).*

Problem 2.2.2. *Prove uniqueness for the Dirichlet and Neumann problems for the reduced Helmholtz equation*

$$\Delta u - ku = 0$$

in a bounded planar domain D, where k is a positive constant.

Problem 2.2.3. *Find the solution $u(x, y)$ of the reduced Helmholtz equation $\Delta u - ku = 0$(k is a positive parameter) in the square $0 < x, y < \pi$, where u satisfies the boundary condition*

$$u(0, y) = 1, u(\pi, y) = u(x, 0) = u(x, \pi) = 0.$$

Problem 2.2.4. *Solve the Laplace equation $\Delta u = 0$ in the square $0 <$*

$x, y < \pi$, *subject to the boundary condition*

$$u(x,0) = u(x,\pi) = 1, u(0,y) = u(\pi,y) = 0.$$

Problem 2.2.5. *Let $u(x,y)$ be a nonconstant harmonic function in the disk $x^2 + y^2 < R^2$. Define for each $0 < r < R$*

$$M(r) = \max_{x^2+y^2=r^2} u(x,y)$$

Prove that $M(r)$ is a monotone increasing function in the interval $(0, R)$

Problem 2.2.6. *Verify that the solution of the Dirichlet problem defined in Example 2.8 is classical.*

Problem 2.2.7. *(a) Compute the Laplace equation in a polar coordinate system. (b) Find a function u, harmonic in the disk $x^2 + y^2 < 6$, and satisfying $u(x,y) = y + y^2$ on the disk's boundary. Write your answer in a Cartesian coordinate system.*

Problem 2.2.8. *(a) Solve the problem*

$$\begin{aligned}
&\Delta u = 0 && 0 < x < \pi, 0 < y < \pi,\\
&u(x,0) = u(x,\pi) = 0 && 0 \le x \le \pi,\\
&u(0,y) = 0 && 0 \le y \le \pi,\\
&u(\pi,y) = \sin y && 0 \le y \le \pi.
\end{aligned}$$

(b) Is there a point $(x,y) \in \{(x,y) \mid 0 < x < \pi, 0 < y < \pi\}$ such that $u(x,y) = 0$?

Problem 2.2.9. *A harmonic function of the form*

$$P_n(x,y) = \sum_{i+j=n} a_{i,j} x^i y^j$$

is called a homogeneous harmonic polynomial of degree n. Denote the space of homogeneous harmonic polynomials of degree n by V_n. What is the dimension of V_n ? Hint Use the polar form of the Laplace equation.

Problem 2.2.10. *Consider the Laplace equation $\Delta u = 0$ in the domain $0 < x, y < \pi$ with the boundary condition*

$$u_y(x, \pi) = x^2 - a, u_x(0, y) = u_x(\pi, y) = u_y(x, 0) = 0.$$

Find all the values of the parameter a for which the problem is solvable. Solve the problem for these values of a.

Problem 2.2.11. *Solve the Laplace equation $\Delta u = 0$ in the domain $x^2+y^2 > 4$, subject to the boundary condition $u(x, y) = y$ on $x^2+y^2 = 4$, and the decay condition $\lim_{|x|+|y|\to\infty} u(x, y) = 0$.*

Problem 2.2.12. *Solve the problem*

$$\begin{aligned} u_{xx} + u_{yy} = 0 \quad & 0 < x < 2\pi, -1 < y < 1, \\ u(x, -1) = 0, \quad u(x, 1) = 1 + \sin 2x \quad & 0 \le x \le 2\pi, \\ u_x(0, y) = u_x(2\pi, y) = 0 \quad & -1 < y < 1. \end{aligned}$$

Problem 2.2.13. *Prove that every nonnegative harmonic function in the disk of radius a satisfies*

Module 3

Integral Equations

3.1 Introduction

An integral equation is an equation in which a function to be determined appears under an integral sign. We consider here principally linear equations, that is, equations in which no nonlinear functions of the unknown function are involved.

Linear integral equations of most frequent occurrence in practice are conventionally divided into two classifications. First, an equation of the form

$$\alpha(x)y(x) = F(x) + \lambda \int_a^b K(x,\xi)y(\xi)d\xi, \tag{3.1}$$

where α, F, and K are given functions and λ, a, and b are con-

stant, is known as a *Fredholm* equation. The function $y(x)$ is to be determined. The given function $K(x, \xi)$, which depends upon the current variable x as well as the auxiliary variable ξ, is known as the *kernel of the integral equation*. If the upper limit of the integral is not a constant, but is identified instead with the current variable, the equation takes the form

$$\alpha(x)y(x) = F(x) + \lambda \int_a^x K(x, \xi)y(\xi)d\xi, \tag{3.2}$$

and is known as a *Volterra equation*.

When $\alpha \not\equiv 0$, the above equations involve the unknown function y both inside and outside the integral. In the special case when $\alpha \equiv 0$, the unknown function appears only under the integral sign, and the equation is known as an integral equation of the *first kind*, while in the case when $\alpha \equiv 1$ the equation is said to be of the *second kind*.

In the more general case when α is not a constant, but is a prescribed function of x, the equation is sometimes called an integral equation of the *third kind*. However, by suitably redefining the unknown function and or the kernel, it is always possible to rewrite such an equation in the form of an equation of the second kind. In particular, when the function $\alpha(x)$ is positive throughout the interval (a, b), equation (3.1) can be rewritten in an equivalent symmetric form

$$\sqrt{\alpha(x)}y(x) = \frac{F(x)}{\sqrt{\alpha(x)}} + \lambda \int_a^b \frac{K(x, \xi)}{\sqrt{\alpha(x)\alpha(\xi)}} \sqrt{\alpha(\xi)}y(\xi)d\xi, \tag{3.3}$$

and hence, in this form, can be considered an integral equation of the second kind in the unknown function $\sqrt{\alpha(x)}y(x)$, with a modified kernel.

Usually, in practice, the functions $\alpha(x), F(x)$, and $K(x,\xi)$ are continuous in (a,b), and it is to be required also that the solution $y(x)$ be continuous in that interval.

In the preceding equations the unknown function depends only upon one independent variable. If an unknown function w depends upon two current variables x and y, the corresponding two-dimensional Fredholm equation is of the form

$$\alpha(x,y)w(x,y) = F(x,y) + \lambda \iint_{\mathscr{R}} K(x,y;\xi,\eta)w(\xi,\eta)d\xi d\eta. \tag{3.4}$$

Certain integral equations can be deduced from or reduced to differential equations. In order to accomplish the reduction, it is frequently necessary to make use of the known formula,

$$\frac{d}{dx}\int_{A(x)}^{B(x)} F(x,\xi)d\xi = \int_A^B \frac{\partial F(x,\xi)}{\partial x}d\xi + F[x,B(x)]\frac{dB}{dx} - F[x,A(x)]\frac{dA}{dx}, \tag{3.5}$$

for differentiation of an integral involving a parameter. The formula of equation (3.5) is valid if both F and $\partial F/\partial x$ are continuous functions of both x and ξ and if both $A'(x)$ and $B'(x)$ are continuous.
As a useful application of this formula, we consider the differentiation of the function $I_n(x)$ defined by the equation

$$I_n(x) = \int_a^x (x-\xi)^{n-1} f(\xi)d\xi, \tag{3.6}$$

where n is a positive integer and a is a constant. With

$$F(x,\xi) = (x-\xi)^{n-1} f(\xi),$$

equation (3.5) gives the derivative of (3.6) in the form

$$\frac{dI_n}{dx} = (n-1)\int_a^x (x-\xi)^{n-2} f(\xi)d\xi + \left[(x-\xi)^{n-1} f(\xi)\right]_{\xi=x}.$$

Hence, if $n > 1$, there follows

$$\frac{dI_n}{dx} = (n-1)I_{n-1} \quad (n>1), \tag{3.7}$$

while if $n = 1$, we have

$$\frac{dI_1}{dx} = f(x). \tag{3.8}$$

Repeated use of (3.7) leads to the general relation

$$\frac{d^k I_n}{dx^k} = (n-1)(n-2)\cdots(n-k)I_{n-k} \quad (n>k) \tag{3.9}$$

In particular, we obtain the result

$$\frac{d^{n-1} I_n}{dx^{n-1}} = (n-1)!\, I_1(x), \tag{3.10}$$

and hence by using (3.8), there follows

$$\frac{d^n I_n}{dx^n} = (n-1)!\, f(x). \tag{3.11}$$

If we notice that $I_n(a) = 0$ when $n \geq 1$, we conclude from (3.9) and

(3.10) that $I_n(x)$ and its first $n-1$ derivatives all vanish when $x = a$.

Thus we may deduce that

$$I_1(x) = \int_a^x f(x_1)\, dx_1,$$

$$I_2(x) = \int_a^x I_1(x_2)\, dx_2 = \int_a^x \int_a^{x_2} f(x_1)\, dx_1 dx_2,$$

and, in the general case,

$$I_n(x) = \int_a^x \int_a^{x_n} \cdots \int_a^{x_3} \int_a^{x_2} f(x_1)\, dx_1 dx_2 \cdots dx_{n-1} dx_n,$$

so that we have the result

$$\int_a^x \int_a^{x_n} \cdots \int_a^{x_3} \int_a^{x_2} f(x_1)\, dx_1 dx_2 \cdots dx_{n-1} dx_n \qquad (3.12)$$
$$= \frac{1}{(n-1)!} \int_a^x (x-\xi)^{n-1} f(\xi) d\xi$$

This relation will be useful in the work that follows. The left-hand member of (3.12) can be interpreted as the result of first integrating f from a to x and then iterating this operation $n-1$ additional times, and is often represented symbolically in the form

$$\overbrace{\int_a^x \cdots \int_a^x}^{n \text{ times}} f(x) \overbrace{dx \cdots dx.}^{n \text{ times}}$$

Exercises

1. (a) If $y''(x) - F(x)$, and y satisfies the initial conditions $y(0) = y_0$ and $y'(0) = y_0'$, show that

$$y(x) = \int_0^x (x - \xi)F(\xi)d\xi + y_0'x + y_0.$$

[Notice that $y'(x) = \int_0^x F(x_1)\,dx_1 + y_0'$ and use (3.12)].

(b) Verify that this expression satisfies the prescribed differential equation and initial conditions.

2. (a) If $y''(x) = F(x)$, and y satisfies the end conditions $y(0) = 0$ and $y(1) = 0$, show that

$$y(x) = \int_0^x (x - \xi)F(\xi)d\xi - x\int_0^1 (1 - \xi)F(\xi)d\xi.$$

(b) Show that the result of part (a) can be written in the form

$$y(x) = \int_0^1 K(x, \xi)F(\xi)d\xi,$$

where $K(x, \xi)$ is defined by the relations

$$K(x,\xi) = \begin{cases} \xi(x-1) & \text{when} \quad \xi < x, \\ x(\xi - 1) & \text{when} \quad \xi > x. \end{cases}$$

(c) Verify directly that the expression obtained satisfies the prescribed differential equation and end conditions.

3.2 Relations between differential and integral equations

We consider first the initial-value problem consisting of the linear second-order differential equation

$$\frac{d^2y}{dx^2} + A(x)\frac{dy}{dx} + B(x)y = f(x) \tag{3.13}$$

together with the prescribed initial conditions

$$y(a) = y_0, \quad y'(a) = y_0'. \tag{3.14}$$

If we solve (3.13) for y'', replace x by x_1, and integrate the result with respect to x_1 over the interval (a, x), using (3.14), there follows

$$y'(x) - y_0' = -\int_a^x A(x_1)\, y'(x_1)\, dx_1 - \int_a^x B(x_1)\, y(x_1)\, dx_1 + \int_a^x f(x_1)\, dx_1$$

or, after integrating the first term on the right by parts,

$$y'(x) = -A(x)y(x) - \int_a^x [B(x_1) - A'(x_1)]\, y(x_1)\, dx_1 + \int_a^x f(x_1)\, dx_1 + A(a)y_0 + y_0'.$$

A second integration then gives the relation

$$y(x) - y_0 = -\int_a^x A(x_1)\, y(x_1)\, dx_1 - \int_a^x \int_a^{x_2} [B(x_1) - A'(x_1)]\, y(x_1)\, dx_1 dx_2 + \int_a^x \int_a^{x_2} f(x_1)\, dx_1 dx_2 + [A(a)y_0 + y_0'](x - a).$$

If use is made of equation (3.12), this equation can be put in the form

$$y(x) = -\int_a^x \{A(\xi) + (x - \xi)[B(\xi) - A'(\xi)]\}\, y(\xi) d\xi + \int_a^x (x - \xi) f(\xi) d\xi + [A(a)y_0 + y_0'](x - a) + y_0$$

or, equivalently,

$$y(x) = \int_a^x K(x, \xi) y(\xi) d\xi + F(x), \tag{3.15}$$

where we have written

$$K(x, \xi) = (\xi - x)[B(\xi) - A'(\xi)] - A(\xi) \tag{3.16}$$

and

$$F(x) = \int_a^x (x - \xi) f(\xi) d\xi + [A(a)y_0 + y_0'](x - a) + y_0 \tag{3.17}$$

This equation is seen to be a *Volterra equation of the second kind.* We may notice that the kernel K is a linear function of the current variable x. It must be assumed, of course, that the coefficients A and B and the function f are such that the indicated integrals exist.

In illustration, the problem

$$\left.\begin{array}{c} \frac{d^2y}{dx^2} + \lambda y = f(x), \\ y(0) = 1, \quad y'(0) = 0 \end{array}\right\} \tag{3.18}$$

is transformed in this way to the integral equation

$$y(x) = \lambda \int_0^x (\xi - x) y(\xi) d\xi + 1 - \int_0^x (\xi - x) f(\xi) d\xi \tag{3.19}$$

Conversely, the use of (3.5) permits the reduction of (3.15) to (3.13) by two differentiations. The initial conditions (3.14) are recovered by setting $x = a$ in (3.15) and in the result of the first differentiation. Thus, differentiation of (3.19) gives

$$\frac{dy}{dx} = -\lambda \int_0^x y(\xi) d\xi + \int_0^x f(\xi) d\xi, \tag{3.20}$$

and a second differentiation leads to the original differential equation. Since the integrals vanish when the upper and lower limits coincide, equations (3.19) and (3.20) supply the initial values $y(0) = 1$ and $y'(0) = 0$.

To illustrate the corresponding procedure in the case of *boundary − value* problems, we consider first a simple example. Starting with the problem

$$\left.\begin{array}{c} \frac{d^2y}{dx^2} + \lambda y = 0, \\ y(0) = 0, \quad y(l) = 0 \end{array}\right\}, \tag{3.21}$$

we obtain after a first integration over $(0, x)$ the relation

$$\frac{dy}{dx} = -\lambda \int_0^x y\,(x_1)\, dx_1 + C, \tag{3.22}$$

where C represents the unknown value of $y'(0)$. A second integration over $(0, x)$ then leads to the relation

$$y(x) = -\lambda \int_0^x (x - \xi) y(\xi) d\xi + Cx. \tag{3.23}$$

While the condition $y(0) = 0$ has been incorporated into this relation, it remains to determine C so that the second end condition $y(l) = 0$ is satisfied. When this condition is imposed on (3.23) there follows

$$\lambda \int_0^l (l - \xi) y(\xi) d\xi = Cl. \tag{3.24}$$

If the value of C so determined is introduced into (20), this relation takes the form

$$y(x) = -\lambda \int_0^x (x - \xi) y(\xi) d\xi + \lambda \frac{x}{l} \int_0^l (l - \xi) y(\xi) d\xi$$

or

$$y(x) = \lambda \int_0^x \frac{\xi}{l} (l - x) y(\xi) d\xi + \lambda \int_x^l \frac{x}{l} (l - \xi) y(\xi) d\xi \tag{3.25}$$

With the abbreviation

$$K(x, \xi) = \begin{cases} \frac{\xi}{l}(l - x) & \text{when} \quad \xi < x \\ \frac{x}{l}(l - \xi) & \text{when} \quad \xi > x \end{cases} \tag{3.26}$$

equation (3.25) becomes

$$y(x) = \lambda \int_0^l K(x, \xi) y(\xi) d\xi. \tag{3.27}$$

Thus, the integral equation corresponding to the *boundary−value* problem (3.21) is a *Fredholm equation of the second kind.*

To recover (3.21) from (3.27), we differentiate the equal members of (3.25) twice, making use of (3.5), as follows:

$$\frac{dy}{dx} = \frac{\lambda}{l}\left[-\int_0^x \xi y(\xi)d\xi + x(l-x)y(x) + \int_x^l (l-\xi)y(\xi)d\xi - x(l-x)y(x)\right]$$
$$= \frac{\lambda}{l}\left[-\int_0^x \xi y(\xi)d\xi + \int_x^l (l-\xi)y(\xi)d\xi\right]$$

and

$$\frac{d^2y}{dx^2} = \frac{\lambda}{l}[-xy(x) - (l-x)y(x)] = -\lambda y(x),$$

in accordance with (3.21). The boundary conditions $y(0) = y(l) = 0$ follow directly from (3.25) by setting $x = 0$ and $x = l$.

We may notice that the kernel (3.26) has different analytic expressions in the two regions $\xi < x$ and $\xi > x$, but that the expressions are equivalent when $\xi = x$. Thus, if we think of K as a *function* of x, for a fixed value of ξ, then K is *continuous* at $x = \xi$. However, the derivative $\partial K/\partial x$ is given by $1 - \xi/l$ when $x < \xi$ and by $-\xi/l$ when $x > \xi$. Thus $\partial K/\partial x$ is discontinuous at $x = \xi$, and it has a finite jump of magnitude — 1 as x *increases* through ξ. Further, we notice that in each region K is a linear function of x, that is, it satisfies the differential equation $\partial^2 K/\partial x^2 = 0$, and K vanishes at the end points $x = 0$ and $x = l$. Finally, it is seen that $K(x, \xi)$ is unchanged if x and ξ are interchanged; that is,

$$K(x, \xi) = K(\xi, x)$$

Kernels having this last property are said to be *symmetric.*

If analogous methods are used in the case of the more general second-order equation

$$\frac{d^2y}{dx^2} + A\frac{dy}{dx} + By = f,$$

with associated end conditions, the kernel so obtained usually is discontinuous at $x = \xi$. However, a kernel which is continuous can be obtained, in general, by a different procedure which is presented in the following section.

Exercises

1. (a) Show that, if $y(x)$ satisfies the differential equation

$$\frac{d^2y}{dx^2} + xy = 1$$

and the conditions $y(0) = y'(0) = 0$, then y also satisfies the Voltera equation

$$y(x) = \int_0^x \xi(\xi - x)y(\xi)d\xi + \frac{1}{2}x^2.$$

(b) Prove that the converse of the preceding statement is also true.

2. Suppose that a sequence of approximate solutions is obtained for the integral equation of above problem , the $(n+1)$th approximation $y^{(n+1)}(x)$ being defined by substitution of the nth approximation into the right-hand member:

$$y^{(n+1)}(x) = \int_0^x \xi(\xi - x)y^{(n)}(\xi)d\xi + \frac{1}{2}x^2.$$

(a) Taking $y^{(0)}(x) = 0$, obtain the functions $y^{(1)}(x) = \frac{1}{2}x^2$ and $y^{(2)}(x) = \frac{1}{2}x^2 - \frac{1}{40}x^5$ as the two succeeding approximations.

(b) Obtain the first two nonvanishing terms in the power series solution of the problem considered in the problem 1(a), in the

form

$$y(x) = a_0 + a_1 x + a_2 x^2 + \cdots,$$

and compare the result with that of part (a).

3.3 The Green's function

We consider first the problem consisting of the differential equation

$$\mathscr{L}y + \Phi(x) = 0, \tag{3.28}$$

where $\mathscr{L}$ is the differential operator

$$\mathscr{L} = \frac{d}{dx}\left(p\frac{d}{dx}\right) + q = p\frac{d^2}{dx^2} + \frac{dp}{dx}\frac{d}{dx} + q, \tag{3.29}$$

together with homogeneous boundary conditions, each of the form

$$\alpha y + \beta\frac{dy}{dx} = 0$$

for some constant values of α and β, which are imposed at the end points of an interval $a \leq x \leq b$. (If $p(x) = 0$ at an end point, the corresponding appropriate end condition may require merely that $y(x)$ remain finite at that point).

The function Φ may be a given direct function of x, or it may also depend upon x indirectly by also involving the unknown function $y(x)$, and so being expressible in the form

$$\Phi(x) = \phi(x, y(x)).$$

In order to obtain a convenient reformulation of this problem, we first attempt the determination of a *Green's function* G which, for a given number ξ, is given by $G_1(x)$ when $x < \xi$ and by $G_2(x)$ when $x > \xi$, and which has the four following properties:

1. The functions G_1 and G_2 satisfy the equation $\mathscr{L}G = 0$ in their intervals of definition; that is, $\mathscr{L}G_1 = 0$ when $x < \xi$, and $\mathscr{L}G_2 = 0$ when $x > \xi$.

2. The function G satisfies the homogeneous conditions prescribed at the end points $x = a$ and $x = b$; that is, G_1 satisfies the condition prescribed at $x = a$, and G_2 that corresponding to $x = b$.

3. The function G is continuous at $x = \xi$; that is, $G_1(\xi) = G_2(\xi)$.

4. The derivative of G has a discontinuity of magnitude $-1/p(\xi)$ at the point $x = \xi$; that is, $G_2'(\xi) - G_1'(\xi) = -1/p(\xi)$.

It is assumed that the function $p(x)$ is continuous and that it differs from zero inside the interval (a, b), so that the discontinuity in the derivative of G is of finite magnitude, and also that $p'(x)$ and $q(x)$ are continuous in (a, b).

We then show that, when the function $G(x, \xi)$ exists, the original formulation of the problem can be transformed to the relation

$$y(x) = \int_a^b G(x, \xi)\Phi(\xi)d\xi, \tag{3.30}$$

in the sense that then (3.30) defines the solution of the problem when Φ

is a given direct function of x, whereas (3.30) constitutes an equivalent integral equation formulation of the problem when Φ involves y.

For the purpose of determining G, let $y = u(x)$ be a nontrivial solution of the associated equation $\mathscr{L}y = 0$ which satisfies the prescribed homogeneous condition at $x = a$, and let $y = v(x)$ be a nontrivial solution of that equation which satisfies the condition prescribed at $x = b$. Then conditions 1 and 2 are satisfied if we write $G_1 = c_1 u(x)$ and $G_2 = c_2 v(x)$, where c_1 and c_2 are constants, so that

$$G = \begin{cases} c_1 u(x) & \text{when} \quad x < \xi, \\ c_2 r(x) & \text{when} \quad x > \xi. \end{cases} \tag{3.31}$$

Conditions 3 and 4 then determine c_1 and c_2 in terms of the value of ξ since condition 3 requires that

$$c_2 v(\xi) - c_1 u(\xi) = 0, \tag{3.32}$$

while condition 4 gives the requirement

$$c_2 v'(\xi) - c_1 u'(\xi) = -\frac{1}{p(\xi)}. \tag{3.33}$$

Equations (3.32, 3.33) possess a unique solution if the determinant does not vanish.

$$W[u(\xi), v(\xi)] - \begin{vmatrix} u(\xi) & v(\xi) \\ u'(\xi) & v'(\xi) \end{vmatrix} = u(\xi)v'(\xi) - v(\xi)u'(\xi) \tag{3.34}$$

This quantity is called the *Wronskian* of the solutions u and v of the equation $\mathscr{L}y = 0$, and it cannot vanish unless the functions u and v are linearly dependent. According to *Abel's formula* this expression has the value $A/p(\xi)$, where A is a certain constant independent of ξ; that is, we have

$$u(\xi)v'(\xi) - v(\xi)u'(\xi) = \frac{A}{p(\xi)} \tag{3.35}$$

With this relation, the solution of (3.32, 3.33) becomes

$$c_1 = -\frac{v(\xi)}{A}, \quad c_2 = -\frac{u(\xi)}{A}, \tag{3.36}$$

and hence (3.31) takes the form

$$G(x,\xi) = \begin{cases} -\frac{1}{A}u(x)v(\xi) & \text{when} \quad x < \xi, \\ -\frac{1}{A}u(\xi)v(x) & \text{when} \quad x > \xi, \end{cases} \tag{3.37}$$

where A is a constant, independent of x and ξ, which is determined by (3.35).

This determination fails if and only if A vanishes, so that u and v are linearly dependent, and hence are each multiples of a certain nontrivial function $U(x)$. In this case, the function $U(x)$ satisfies the equation $\mathscr{L}y = 0$ and both end conditions. Thus, for example, since the function $U(x) = 1$ solves the problem $d^2y/dx^2 = 0, y'(0) = y'(1) = 0$, the Green's function does not exist for the expression $\mathscr{L}y = d^2y/dx^2$, relevant to the end conditions $y'(0) = -y'(1) = 0$.

We now show that, with the definition of equation (3.37), the relation

$$y(x) = \int_a^b G(x,\xi)\Phi(\xi)d\xi \tag{3.38}$$

implies the differential equation

$$\mathscr{L}y + \Phi(x) = 0, \tag{3.39}$$

together with the prescribed boundary conditions. For this purpose, we write (3.38) in the explicit form

$$y(x) = -\frac{1}{A}\left[\int_a^x v(x)u(\xi)\Phi(\xi)d\xi + \int_x^b u(x)v(\xi)\Phi(\xi)d\xi\right]. \tag{3.40}$$

Two differentiation, then lead to the relations

$$y'(x) = -\frac{1}{A}\left[\int_a^x v'(x)u(\xi)\Phi(\xi)d\xi + \int_x^b u'(x)v(\xi)\Phi(\xi)d\xi\right] \tag{3.41}$$

and

$$\begin{aligned} y''(x) &\\ &= -\frac{1}{A}\left[\int_a^x v''(x)u(\xi)\Phi(\xi)d\xi + \int_x^b u''(x)v(\xi)\Phi(\xi)d\xi\right] \\ &\quad - \frac{1}{A}\left[v'(x)u(x) - u'(x)v(x)\right]\Phi(x). \end{aligned} \tag{3.42}$$

If we form the combination

$$\mathscr{L}y \equiv p(x)y''(x) + p'(x)y'(x) + q(x)y(x)$$

from these results, and make use of (3.35), there follows

$$\begin{aligned}\mathscr{L}y(x) &= -\frac{1}{A}\left\{\int_a^x [\mathscr{L}v(x)]u(\xi)\Phi(\xi)d\xi + \int_x^b [\mathscr{L}u(x)]v(\xi)\Phi(\xi)d\xi\right\} \\ &\quad - \frac{1}{A}\left[p(x)\cdot\frac{A}{p(x)}\cdot\Phi(x)\right]\end{aligned} \tag{3.43}$$

But since $u(x)$ and $v(x)$ satisfy $\mathscr{L}y = 0$, the two integrands vanish identically, and this relation becomes merely

$$\mathscr{L}y(x) = -\Phi(x), \tag{3.44}$$

so that (3.38) implies (3.39). That is, a function y satisfying (3.38) satisfies the differential equation (3.39). Also, since (3.40) and (3.41) give

$$y(a) = -\frac{1}{A}u(a)\int_a^b v(\xi)\Phi(\xi)d\xi \tag{3.45}$$

$$y'(a) = -\frac{1}{A}u'(a)\int_a^b v(\xi)\Phi(\xi)d\xi \tag{3.46}$$

it follows that a function y defined by (3.38) satisfies the same homogeneous condition at $x = a$ as the function u. But this condition was specified as that which is imposed on the solution of (3.39) at $x = a$. A similar statement applies to the satisfaction of the condition prescribed at $x = b$.

Conversely, the differential equation (3.39) together with the associated end conditions can be shown to imply the relation (3.38), so that the two formulations are entirely equivalent.

In particular, by identifying $\Phi(x)$ with $\lambda r(x)y(x) - f(x)$ we may deduce the equivalence of the differential equation

$$\mathscr{L}y(x) + \lambda r(x)y(x) = f(x) \tag{3.47}$$

with associated homogeneous conditions imposed at the ends of the interval (a, b), and the Fredholm integral equation

$$y(x) = \lambda \int_a^b G(x,\xi)r(\xi)y(\xi)d\xi - \int_a^b G(x,\xi)f(\xi)d\xi \tag{3.48}$$

where G is the relevant Green's function.

We may notice that the kernel $K(x,\xi)$ of (3.48) is actually the product $G(x,\xi)r(\xi)$. While the definition (3.37) shows that $G(x,\xi)$ is symmetric, the product $K(x,\xi)$ is not symmetric unless $r(x)$ is a constant. However, if we write

$$\sqrt{r(x)}y(x) = Y(x), \tag{3.49}$$

under the assumption that $r(x)$ is nonnegative over (a, b), as is usually the case in practice, equation (3.48) can be written in the form

$$Y(x) = \lambda \int_a^b \widetilde{K}(x,\xi) Y(\xi) d\xi - \int_a^b \tilde{K}(x,\xi) \frac{f(\xi)}{\sqrt{r(\xi)}} d\xi, \tag{3.50}$$

where $\widetilde{K}$ is defined by the relation

$$\widetilde{K}(x,\xi) = \sqrt{r(x)r(\xi)} G(x,\xi), \tag{3.51}$$

and hence possesses the same symmetry as G. The importance of symmetry will be seen in later considerations.

In the special case when the operator $\mathscr{L}$ and the associated end conditions are such that

$$\mathscr{L}y = y'', \quad y(0) = y(l) = 0, \tag{3.52}$$

it is readily verified that the relevant Green's function G is identified with the kernel K defined by,

$$G(x,\xi) = \begin{cases} \frac{x}{l}(l-\xi) & (x < \xi), \\ \frac{\xi}{l}(l-x) & (x > \xi). \end{cases} \tag{3.53}$$

Thus, in particular, the solution of the problem

$$y'' = f(x), \quad y(0) = y(l) = 0 \tag{3.54}$$

is

$$y(x) = -\int_0^l G(x,\xi) f(\xi) d\xi, \tag{3.55}$$

whereas the problem

$$y'' + \lambda r y = f(x), \quad y(0) = y(l) = 0 \tag{3.56}$$

is equivalent to the integral equation

$$y(x) = \lambda \int_0^l G(x,\xi) r(\xi) y(\xi) d\xi - \int_0^l G(x,\xi) f(\xi) d\xi \tag{3.57}$$

When the prescribed end conditions are not homogeneous, a modified procedure is needed. In this case, we denote by $G(x,\xi)$ the Green's function corresponding to the associated *homogeneous* end conditions, and attempt to determine a function $P(x)$ such that the relation

$$y(x) = P(x) + \int_a^b G(x,\xi)\Phi(\xi) d\xi \tag{3.58}$$

is equivalent to the differential equation

$$\mathscr{L} y(x) + \Phi(x) = 0, \tag{3.59}$$

together with the prescribed non-homogeneous end conditions. Since

$$\mathscr{L} \int_a^b G(x,\xi)\Phi(\xi) d\xi = -\Phi(x),$$

the requirement that (3.58) imply (3.59) takes the form

$$\mathscr{L}P(x) = 0 \tag{3.60}$$

and, since the second term in (3.58) satisfies the associated homogeneous end conditions, it follows that the function $P(x)$ in (3.58) must be the solution of (3.60) which satisfies the prescribed non-homogeneous end conditions. The existence of $P(x)$ is insured when $G(x, \xi)$ itself exists. The equivalence of the two formulations follows from the previously established results when the new unknown function $w(x) = y(x) - P(x)$ is introduced.

In illustration, in order to transform the problem

$$y'' + xy = 1, \quad y(0) = 0, \quad y(l) = 1 \tag{3.61}$$

to an integral equation, we again use the Green's function (3.53), corresponding to the associated homogeneous conditions $y(0) = y(l) = 0$, and determine $P(x)$ such that $P''(x) = 0, P(0) = 0, P(l) = 1$, and hence $P(x) = x/l$. With $\Phi(x) = xy(x) - 1$, equation (3.58) then becomes

$$y(x) = \frac{x}{l} + \int_0^l G(x, \xi)[\xi y(\xi) - 1]d\xi$$

and reduces to the form

$$y(x) = \frac{x}{2l}\left(1 + \frac{x}{l}\right) + \int_0^l G(x, \xi)\xi y(\xi)d\xi. \tag{3.62}$$

As an explicit illustration of the use of (3.37), we consider the problem

$$\left.\begin{array}{r} x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(\lambda x^2 - 1\right) y = 0, \\ y(0) = 0, \quad y(1) = 0 \end{array}\right\}. \tag{3.63}$$

The differential equation is first put into the form of (3.28),

$$\frac{d}{dx}\left(x\frac{dy}{dx}\right) + \left(-\frac{1}{x} + \lambda x\right) y = 0,$$

from which there follows

$$\mathscr{L}y = \frac{d}{dx}\left(x\frac{dy}{dx}\right) - \frac{y}{x}, \quad p = x, \quad q = -\frac{1}{x}, \quad r = x. \tag{3.64}$$

The general solution of the equation $\mathscr{L}y = 0$ is found to be

$$y = c_1 x + c_2 x^{-1}.$$

As a solution for which $y(0) = 0$ we may take $y = u(x)$, where

$$u(x) = x, \tag{3.65}$$

and as a solution for which $y(1) = 0$ we may take $y = v(x)$, where

$$v(x) = \frac{1}{x} - x. \tag{3.66}$$

The Wronskian of u and v is then given by

$$u(x)v'(x) - v(x)u'(x) = -\frac{2}{x} \equiv \frac{-2}{p(x)},$$

and hence, with the notation of (3.35), we have

$$A = -2. \tag{3.67}$$

Thus (3.37) becomes

$$G(x,\xi) = \begin{cases} \frac{x}{2\xi}\left(1-\xi^2\right) & \text{when} \quad x < \xi, \\ \frac{\xi}{2x}\left(1-x^2\right) & \text{when} \quad x > \xi. \end{cases} \tag{3.68}$$

It follows from (3.48) that the problem (3.63) then corresponds to the integral equation

$$y(x) = \lambda \int_0^1 G(x,\xi)\xi y(\xi) d\xi. \tag{3.69}$$

It is easily seen that the Bessel equation (3.63) has no solution other than the trivial solution $y \equiv 0$, satisfying the prescribed end conditions, unless λ satisfies the characteristic equation

$$J_1\left(\sqrt{\lambda_n}\right) = 0, \tag{3.70}$$

in which case the solution is

$$y = cJ_1\left(\sqrt{\lambda_n}x\right), \tag{3.71}$$

where c is arbitrary. The same statement must then apply to the integral

equation (3.69), with G given by (3.68).

If the condition $y(1) = 0$ were replaced by the condition $y(1) = k$ in (3.63), the function $P(x) = kx$ would be added to the right-hand member of (3.69). Because of the singular point of the differential equation at $x = 0$, a nonzero value of $y(0)$ could not be prescribed.

A completely analogous procedure can be used in transforming a boundary-value problem consisting of a differential equation of order n, and relevant homogeneous boundary conditions, to a Fredholm integral equation. The Green's function corresponding to $\mathscr{L}y$ in the interval (a, b) then is to possess the following properties:

1. G satisfies the equation $\mathscr{L}G = 0$ when $x < \xi$ and when $x > \xi$.

2. G satisfies the prescribed homogeneous boundary conditions.

3. G and its first $n - 2x$-derivatives are continuous at $x = \xi$.

4. The $(n-1)$ th x-derivative of G has a jump of magnitude $-1/s(\xi)$ as x increases through ξ, where $s(x)$ is the coefficient of d^n/dx^n in $\mathscr{L}$.

With the function G so defined, the differential cquation $\mathscr{L}y+\Phi(x) = 0$, subject to the relevant homogeneous end conditions, corresponds to the relation

$$y(x) = \int_a^b G(x, \xi)\Phi(\xi)d\xi$$

and, in particular, the problem consisting of the equation $\mathscr{L}y + \lambda ry = f$ and the prescribed homogeneous end conditions is equivalent to the Fredholm equation

$$y(x) = \lambda \int_a^b G(x,\xi) r(\xi) y(\xi) d\xi - \int_a^b G(x,\xi) f(\xi) d\xi$$

In addition, for those fourth-order operators which are expressed in the $self-adjoint$ form

$$\mathscr{L} = \frac{d^2}{dx^2}\left[s(x)\frac{d^2}{dx^2}\right] + \frac{d}{dx}\left[p(x)\frac{d}{dx}\right] + q(x), \tag{3.72}$$

it will be found that the Green's function $G(x,\xi)$ is symmetric. Most of the linear fourth-order operators occurring in practice can be expressed in this form.

Exercises

1. Transform the problem

$$\frac{d^2y}{dx^2} + xy = 1, \quad y(0) - y(1) = 0$$

to the integral equation

$$y(x) = \int_0^1 G(x,\xi)\xi y(\xi) d\xi - \frac{1}{2}x(1-x),$$

where $G(x,\xi) = x(1-\xi)$ when $x < \xi$ and $G(x,\xi) = \xi(1-x)$ when $x > \xi$.

2. Transform the problem

$$\frac{d^2y}{dx^2} + y + \epsilon y^2 = f(x), \quad y(0) = 0, \quad y'(1) = 0$$

to a nonlinear Fredholm integral equation in each of the two fol-

lowing ways:

(a) Use the method of Section 3.2 with $\mathscr{L}y = y^*$.

(b) Proceed as in part (a) with $\mathscr{L}y = y^* + y$.

3. Transform the problem

$$\frac{d^2y}{dx^2} + y = x, \quad y(0) = 0, \quad y'(1) = 0$$

to a Fredholm integral equation.

4. Transform the problem

$$\frac{d^2y}{dx^2} + y = x, \quad y(0) = 1, \quad y'(1) = 0$$

to a Fredholm integral equation.

5. (a) Determine $p(x)$ and $q(x)$ in such a way that the equation

$$x^2\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + 2y = 0$$

is equivalent to the equation $\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy = 0$, thus showing that the equation can be written in the "self-adjoint" form

$$\frac{d}{dx}\left(\frac{1}{x^2}\frac{dy}{dx}\right) + \frac{2}{x^4}y = 0.$$

(b) Verify that $u(x) = x$ and $v(x) = x^2$ are linearly independent solutions of the equation of part (a).

(c) Show that, if we take $u(x) = a_1x + a_2x^2$ and $v(x) = b_1x + b_2x^2$, there follows $A = a_1b_2 - a_2b_1$. [Notice that hence $A \neq 0$ unless

u and v are in a constant ratio, and hence are not linearly independent.]

6. (a) Show that the Green's function for the Bessel operator of order n,

$$\mathscr{L}y = \frac{d}{dx}\left(x\frac{dy}{dx}\right) - \frac{n^2}{x}y,$$

relevant to the end conditions $y(0) = y(1) = 0$, is of the form

$$G(x,\xi) = \begin{cases} \frac{x^n/\xi^n}{2n}\left(1-\xi^{2n}\right) & \text{when} \quad x < \xi, \\ \frac{\xi^n/x^n}{2n}\left(1-x^{2n}\right) & \text{when} \quad x > \xi, \end{cases}$$

when $n \neq 0$.

(b) Use the result of part (a) to reduce the problem

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(\lambda x^2 - n^2\right)y = 0, \quad y(0) = y(1) = 0$$

to an integral equation, when $n \neq 0$.

7. Transform the problem

$$\frac{d^4y}{dx^4} + \Phi(x) = 0, \quad y(0) = y'(0) = y(1) = y'(1) = 0$$

to the relation

$$y(x) = \int_0^1 G(x,\xi)\Phi(\xi)d\xi,$$

where

$$G(x,\xi) = \begin{cases} \frac{1}{6}x^2(1-\xi)^2(2x\xi + x - 3\xi) & \text{when} \quad x < \xi, \\ \frac{1}{8}\xi^2(1-x)^2(2x\xi + \xi - 3x) & \text{when} \quad x > \xi. \end{cases}$$

8. (a) If $\mathscr{S}$ is the self-adjoint operator $\mathscr{L} = \frac{d}{dx}\left(p\frac{dy}{dx}\right) + q$, show that

$$\int_A^B f(x)[\mathscr{L}g(x)]dx$$
$$= \int_A^B g(x)[\mathscr{L}f(x)]dx + [p(x)\{f(x)g'(x) - f'(x)g(x)\}]_A^B \tag{3.73}$$

for any functions $f(x)$ and $g(x)$ which possess a continuous first derivative in (A, B).

[Write the left-hand member in the form $\int_A^B f\,(pg')'\,dx + \int_A^B fqgdx$, and integrate the first term by parts. Then integrate the resultant term again by parts, in an appropriate way. This formula is known as Green's formula for the operator $\mathscr{L}$.]

(b) If $f(x)$ and $g(x)$ satisfy identical homogeneous conditions at the ends $x = A$ and $x = B$ of the interval, show that the integrated terms vanish, and there follows

$$\int_A^B f(x)[\mathscr{L}g(x)]dx = \int_A^B g(x)[\mathscr{L}f(x)]dx$$

[Notice that this situation exists if at each end we require the vanishing of each function, or of the derivative of each function, or of a linear combination of each function and its derivative. Further, if $p(x)$ vanishes at an end, we may require at that end only that each function and its derivative remain finite. Also, if $p(B) = p(A)$, we may require merely that each function and its derivative take on the same values at $x = B$ as at $x - A$.]

3.4 Fredholm equations with separable kernels

We shall speak of a kernel $K(x, \xi)$ as *separable* if it can be expressed as the sum of a finite number of terms, each of which is the product of a function of x alone and a function of ξ alone. Such a kernel is thus expressible in the form

$$K(x, \xi) = \sum_{n=1}^{N} f_n(x) g_n(\xi). \tag{3.74}$$

There is no loss in generality if we assume that the N functions $f_n(x)$ are linearly independent in the relevant interval.

Any polynomial in x and ξ is of this type. Further, we may notice, for example, that the kernel $\sin(x + \xi)$ is separable in this sense, since we can write

$$\sin(x + \xi) = \sin x \cos \xi + \cos x \sin \xi.$$

Integral equations with separable kernels do not occur frequently in practice. However, they are easily treated and, furthermore, the results of their consideration lead to a better understanding of integral equations of more general type. Also, it is often possible to apply the methods to be developed in this section to the approximate solution of Fredholm equations in which the kernel can be satisfactorily approximated by a polynomial in x and ξ, or by a separable kernel of more general form.

A Fredholm equation of the second kind, with (3.74) as its kernel, can be written in the form

$$y(x) = \lambda \int_a^b K(x,\xi) y(\xi) d\xi + F(x) \tag{3.75}$$

or

$$y(x) = \lambda \int_a^b \left[\sum_{n=1}^{N} f_n(x) g_n(\xi) \right] y(\xi) d\xi + F(x)$$

or

$$y(x) = \lambda \sum_{n=1}^{N} f_n(x) \left[\int_a^b g_n(\xi) y(\xi) d\xi \right] + F(x) \tag{3.76}$$

It is seen that the coefficients of $f_1(\mathrm{x}), f_2(\mathrm{x}), \cdots, f_{.}(x)$ in (3.76) are constants, although their values are unknown. If we introduce the abbreviation

$$c_n = \int_a^b g_n(x) y(x) dx \quad (n = 1, 2, \cdots, N), \tag{3.77}$$

equation (3.76) takes the form

$$y(x) = F(x) + \lambda \sum_{n=1}^{N} c_n f_n(x). \tag{3.78}$$

This is thus the form of the required solution of the integral equation (3.75), and it remains only to determine the N constants $c_1, c_2, \cdots, c_N$.

For this purpose, we may obtain N equations involving only these N constants and calculable quantities by multiplying both members of (3.78) successively by $g_1(x), g_2(x), \cdots, g_{.v}(x)$, and integrating the results over the interval (a, b). This procedure introduces calculable integrals which may be denoted as follows:

$$\alpha_{mn} = \int_a^b g_m(x) f_n(x) dx, \quad \beta_m = \int_a^b g_m(x) F(x) dx \tag{3.79}$$

With these abbreviations, the N equations so obtained can be written in the form

$$\left.\begin{array}{c} (1-\lambda\alpha_{11})\, c_1 - \lambda\alpha_{12} c_2 - \lambda\alpha_{13} c_3 + \cdots - \lambda\alpha_{1N} c_N = \beta_1 \\ -\lambda\alpha_{21} c_1 + (1-\lambda\alpha_{22})\, c_2 - \lambda\alpha_{23} c_3 + \cdots - \lambda\alpha_{2N} c_N = \beta_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \alpha_{N1} c_1 - \lambda\alpha_{N2} c_2 - \lambda\alpha_{N3} c_3 + \cdots + (1-\lambda\alpha_{NN})\, c_N = \beta_N \end{array}\right\} \tag{3.80}$$

This set of equations possesses a unique solution for the c 's if and only if the determinant Δ of the coefficients of the c's does not vanish. In the matrix notation, the set may be written in the abbreviated form

$$(\mathbf{I} - \lambda\mathbf{A})\mathbf{c} = \beta, \tag{3.81}$$

where I is the unit matrix of order N and A is the matrix $[\alpha_{ij}]$.

If the function $F(x)$ is identically zero in (3.75), the integral equation is said to be homogeneous, and is obviously satisfied by the trivial solution $y(x) \equiv 0$, corresponding to the trivial solution $c_1 = c_2 = \cdots = c_N = 0$ of (3.80) when the right-hand members vanish. Unless the determinant $\Delta \equiv |\mathbf{I} - \lambda\mathbf{A}|$ vanishes, this is the only solution. However, if $\Delta = 0$, at least one of the c's can be assigned arbitrarily, and the remaining c's can be determined accordingly. Thus, in such cases, infinitely many solutions of the integral equation (3.75) exist.

Those values of λ for which $\Delta(\lambda) = 0$ are known as the characteristic

values (or eigenvalues), and any nontrivial solution of the homogeneous integral equation (with a convenient choice of the arbitrary constant or constants) is then called a corresponding characteristic function (eigenfunction) of the integral equation. If k of the constants $c_1, c_2, \cdots, c_N$ can be assigned arbitrarily for a given characteristic value of λ, then k linearly independent corresponding characteristic functions are obtained.

If the function $F(x)$ is not identically zero, but is orthogonal to all the functions $g_1(x), g_2(x), \cdots, g_N(x)$, equation (3.79) shows that the right-hand members of (3.80) again vanish. The preceding discussion again applies to this case, except for the fact that here the solution (3.78) of the integral equation involves also the function $F(x)$. The trivial values $c_1 = c_2 = \cdots = c_N = 0$ thus lead to the solution $y = F(x)$. Solutions corresponding to characteristic values of λ are now expressed as the sum of $F(x)$ and arbitrary multiples of characteristic functions.

Finally, if at least one right-hand member of (3.80) does not vanish, a unique nontrivial solution of (3.80) exists, leading to a unique nontrivial solution of the integral equation (3.75), if the determinant $\Delta(\lambda)$ does not vanish. However, in this case, if $\Delta(\lambda)$ does vanish equations (3.80) are either incompatible, and no solution exists, or they are redundant, and infinitely many solutions exist.

Exercises

1. (a) Show that the characteristic values of λ for the equation

$$y(x) = \lambda \int_0^{2\pi} \sin(x + \xi) y(\xi) d\xi$$

are $\lambda_1 = 1/\pi$ and $\lambda_3 = -1/\pi$, with corresponding characteristic functions of the form $y_1(x) = \sin x + \cos x$ and $y_2(x) = \sin x - \cos x$.

(b) Obtain the most general solution of the equation

$$y(x) = \lambda \int_0^{2\pi} \sin(x + \xi) y(\xi) d\xi + F(x)$$

when $F(x) = x$ and when $F(x) = 1$, under the assumption that $\lambda \neq \pm 1/\pi$.

(c) Prove that the equation

$$y(x) = \frac{1}{\pi} \int_0^{2\pi} \sin(x + \xi) y(\xi) d\xi + F(x)$$

possesses no solution when $F(x) = x$, but that it possesses infinitely many solutions when $F(x) = 1$. Determine all such solutions.

(d) Determine the most general form of the prescribed function $F(x)$, for which the integral equation

$$\int_0^{2\pi} \sin(x + \xi) y(\xi) d\xi = F(x)$$

of the first kind, possesses a solution.

2. Consider the equation

$$y(x) = F(x) + \lambda \int_0^{2x} \cos(x + \xi) y(\xi) d\xi$$

(a) Determine the characteristic values of λ and the corresponding characteristic functions.

(b) Express the solution in the form

$$y(x) = F(x) + \lambda \int_0^{2x} \Gamma(x, \xi; \lambda) F(\xi) d\xi$$

when λ is not characteristic.

(c) Obtain the general solution (when it exists) if $F(x) = \sin x$, considering all possible cases.

3.5 Illustrative example

The several possible cases just discussed may be illustrated by a consideration of the integral equation

$$y(x) = \lambda \int_0^1 (1 - 3x\xi) y(\xi) d\xi + F(x) \tag{3.82}$$

This equation can be rewritten in the form

$$y(x) = \lambda (c_1 - 3c_2 x) + F(x) \tag{3.83}$$

where

$$c_1 = \int_0^1 y(\xi) d\xi, \quad c_2 = \int_0^1 \xi y(\xi) d\xi \tag{3.84}$$

To determine c_1 and c_2, we multiply both sides of (3.83) successively by 1 and x and integrate the results over $(0, 1)$, to obtain the equations

$$c_1 = \lambda \left(c_1 - \frac{3}{2} c_2 \right) + \int_0^1 F(x) dx,$$
$$c_2 = \lambda \left(\frac{1}{2} c_1 - c_2 \right) + \int_0^1 x F(x) dx,$$

or

$$\left.\begin{array}{l}(1-\lambda)c_1+\frac{3}{2}\lambda c_2=\int_0^1 F(x)dx\\ \frac{1}{2}\lambda c_1+(1+\lambda)c_2=\int_0^1 xF(x)dx\end{array}\right\} \tag{3.85}$$

The determinant of coefficients is given by

$$\Delta(\lambda)=\left|\begin{array}{cc}1-\lambda & \frac{3}{2}\lambda\\ -\frac{1}{2}\lambda & 1+\lambda\end{array}\right|=\frac{1}{4}\left(4-\lambda^2\right). \tag{3.86}$$

It follows that a unique solution exists if and only if

$$\lambda\neq\pm 2, \tag{3.87}$$

and is obtained by solving (3.85) for c_1 and c_2, and introducing the results into (3.83). In particular, if $F(x)=0$ and $\lambda\neq\pm 2$, the only solution is the trivial one, $y(x)=0$. The numbers $\lambda=\pm 2$ are the characteristic numbers of the problem.

If $\lambda=+2$, equations (3.85) take the form

$$\left.\begin{array}{l}-c_1+3c_2=\int_0^1 F(x)dx,\\ -c_1+3c_2=\int_0^1 xF(x)dx\end{array}\right\}, \tag{3.88}$$

while if $\lambda=-2$ equations (3.85) become

$$\left.\begin{array}{l}c_1-c_2=\frac{1}{3}\int_0^1 F(x)dx,\\ c_1-c_2=\int_0^1 xF(x)dx\end{array}\right\}. \tag{3.89}$$

Equations (3.88) are incompatible unless the prescribed function $F(x)$ satisfies the condition

$$\int_0^1 F(x)dx=\int_0^1 xF(x)dx \quad \text{or} \quad \int_0^1 (1-x)F(x)dx=0 \tag{3.90}$$

whereas (3.89) are incompatible unless

$$\frac{1}{3}\int_0^1 F(x)dx = \int_0^1 xF(x)dx \text{ or } \int_0^1 (1-3x)F(x)dx = 0, \qquad (3.91)$$

in which cases the corresponding equation pairs (3.88) or (3.89) are redundant.

We consider first the case when

$$F(x) = 0,$$

so that (3.82) is *homogeneous*. Then, if $\lambda \neq \pm 2$, the only solution is the trivial one $y(x) = 0$, as was mentioned above. If $\lambda = 2$ (and $F = 0$) equations (3.88) are redundant, and either equation gives the single condition $c_1 = 3c_2$. Thus (3.83) then gives the solution

$$y(x) = A(1-x) \text{ when } \lambda = 2, \qquad (3.92)$$

where $A = 6c_2$ is an *arbitrary* constant. Thus the function $1 - x$ (or any convenient nonzero multiple of that function) is the *characteristic function* corresponding to the characteristic number $\lambda = +2$. In a similar way, we find the solution

$$y(x) = B(1-3x) \quad \text{when} \quad \lambda = -2 \qquad (3.93)$$

where $B = -2c_1 = -2c_2$ is an arbitrary constant, so that $1 - 3x$ is the characteristic function corresponding to $\lambda = -2$.

Equation (3.83) shows that any solution of (3.82) is expressible in the form

$$y(x) = F(x) + C_1(1-x) + C_2(1-3x), \qquad (3.94)$$

where we have written $C_1 = 3\lambda(c_1 - c_2)/2$ and $C_2 = \lambda(3c_2 - c_1)/2$. Thus it follows that any *solution* of (3.82) can be expressed as the sum of $F(x)$ and some linear combination of the characteristic functions. The fact that this statement can be applied to a wide class of integral equations is of basic importance, as will be seen.

In the *nonhomogeneous* case, $F(x) \not\equiv 0$, a unique solution exists if $\lambda \neq \pm 2$. If $\lambda = 2$, equation (3.90) shows that no solution exists unless $F(x)$ is orthogonal to $1 - x$ over the relevant interval $(0, 1)$, that is, unless $F(x)$ is orthogonal to the characteristic function corresponding to $\lambda = 2$. If F satisfies this restriction equations (3.88) are again equivalent. Hence, if we use (3.88), we may obtain $c_1 = 3c_2 - \int_0^1 F(x)dx$, so that (3.83) gives the solution as follows:

$$\lambda = 2: \quad y(x) = F(x) - 2\int_0^1 F(x)dx + A(1 - x),$$

when

$$\int_0^1 (1 - x)F(x)dx = 0. \tag{3.95}$$

Here $A = 6c_2$ is again an arbitrary constant. Thus, in this case, *infinitely many solutions exist, differing from each other by a multiple of the relevant characteristic function.*

Similarly, if $\lambda = -2$ there is no solution unless $F(x)$ is orthogonal to $1 - 3x$ over $(0, 1)$, in which case infinitely many solutions exist as follows:

$$\lambda = -2: \quad y(x) = F(x) - \frac{2}{3}\int_0^1 F(x)dx + B(1 - 3x)$$

when

$$\int_0^1 (1 - 3x)F(x)dx = 0. \tag{3.96}$$

Here $B = -2c_2$ is an arbitrary constant.

3.6 Hilbert-Schmidt theory

In those cases when the kernel $K(x, \xi)$ of a homogeneous Fredholm equation is not of the form (3.74), in particular, if $K(x, \xi)$ is given by different analytic expressions in the intervals for which $x < \xi$ and $x > \xi$, there are generally infinitely many characteristic numbers $\lambda_n (n = 1, 2, 3, \cdots)$, each corresponding to a characteristic function defined within an arbitrary multiplicative constant. In exceptional cases, a given characteristic number λ_k may correspond to two or more independent characteristic functions. In this section we investigate certain properties of these characteristic functions.

Let $y_m(x)$ and $y_n(x)$ be characteristic functions corresponding respectively to two *different* characteristic numbers λ_m and λ_n of the homogeneous Fredholm equation

$$y(x) = \lambda \int_a^b K(x, \xi) y(\xi) d\xi \tag{3.97}$$

and *suppose* that the kernel $K(x, \xi)$ is *symmetric*, so that

$$K(x, \xi) = K(\xi, x) \tag{3.98}$$

As has been indicated in preceding sections, such kernels are of frequent occurrence in the formulation of physically motivated problems. We may notice that $\lambda = 0$ cannot be a characteristic number since it leads necessarily to the trivial solution $y(x) = 0$.

The functions y_m and y_n accordingly must satisfy the equations

$$y_m(x) = \lambda_m \int_a^b K(x,\xi) y_m(\xi) d\xi, \tag{3.99}$$

$$y_n(x) = \lambda_n \int_a^b K(x,\xi) y_n(\xi) d\xi. \tag{3.100}$$

If we multiply both members of (3.99) by $y_n(x)$, and integrate the results with respect to x over (a,b), there then follows

$$\int_a^b y_m(x) y_n(x) dx = \lambda_m \int_a^b y_n(x) \left[\int_a^b K(x,\xi) y_m(\xi) d\xi \right] dx. \tag{3.101}$$

If the order of integration is reversed in the right-hand member, equation (3.101) becomes

$$\int_a^b y_m(x) y_n(x) dx = \lambda_m \int_a^b y_m(\xi) \left[\int_a^b K(x,\xi) y_n(x) dx \right] d\xi. \tag{3.102}$$

We now make use of the assumed *symmetry* (3.98) to rewrite the inner integral on the right in the form

$$\int_a^b K(\xi,x) y_n(x) dx$$

But this integral differs from the coefficient of λ_n in (3.100) only in that x and ξ are interchanged, and hence it is equivalent to

$$\frac{1}{\lambda_n} y_n(\xi),$$

so that (3.102) becomes

$$\int_a^b y_m(x)y_n(x)dx = \frac{\lambda_m}{\lambda_n}\int_a^b y_m(\xi)y_n(\xi)d\xi. \tag{3.103}$$

Since the integrals in (3.103) are equivalent, (3.103) can be rewritten in the form

$$(\lambda_m - \lambda_n)\int_a^b y_m(x)y_n(x)dx = 0. \tag{3.104}$$

Thus we conclude that if $y_m(x)$ and $y_n(x)$ are *characteristic functions* of (3.97) corresponding to distinct characteristic numbers, then $y_m(x)$ and $y_n(x)$ are *orthogonal over the interval* (a, b).

If two or more linearly independent characteristic functions correspond to the same characteristic number, then an equal number of orthogonalized linear combinations can be formed by the Gram-Schmidt procedure. In the remainder of this chapter it will be assumed that this has been done, when such exceptional cases arise.

It is important to notice that the preceding results apply only to a *symmetric kernel*.

We show next that the *characteristic numbers* of a *Fredholm equation* with a *real symmetric kernel* are all *real*. This result is established by noticing that if λ_m were an imaginary characteristic number, corresponding to a complex characteristic function $y_m(x)$, then the complex conjugate number $\bar{\lambda}_m$ would necessarily also be a characteristic number, corresponding to the characteristic function $\bar{y}_m(x)$ which is the complex conjugate of $y_m(x)$. Hence, by replacing λ_n by $\bar{\lambda}_m$ and y_n by $\bar{y}_m$ in (3.104), it would follow that

$$\left(\lambda_m - \bar{\lambda}_m\right)\int_a^b y_m(x)\bar{y}_m(x)dx = 0. \tag{3.105}$$

If we write $\lambda_m = \alpha_m + i\beta_m$ and $y_m(x) = f_m(x) + ig_m(x)$, this relation takes the form

$$2i\beta_m \int_a^b \left(f_m^2 + g_m^2\right) dx = 0 \tag{3.106}$$

But, since $y_m(x) \not\equiv 0$, the integral cannot vanish, and we conclude that the imaginary part of λ_m must vanish, as was to be shown.

A Fredholm equation with a nonsymmetric kernel may possess characteristic numbers which are not real.

In more advanced works the following basic theorem is established:

Any function $f(x)$ which can be generated from a continuous function $\Phi(x)$ by the operation $\int_a^b K(x,\xi)\Phi(\xi)d\xi$, where $K(x,\xi)$ is continuous, real, and symmetric, so that

$$f(x) = \int_a^b K(x,\xi)\Phi(\xi)d\xi$$

for some continuous function Φ, can be represented over (a,b) by a linear combination of the characteristic functions $y_1(x), y_2(x), \cdots$, of the homogeneous Fredholm integral equation (3.97) with $K(x,\xi)$ as its kernel.

Because of the orthogonality, the coefficients in the representation

$$f(x) = \sum_n A_n y_n(x) \quad (a \le x \le b) \tag{3.107}$$

are then determined by the familiar formula

$$A_n \int_a^b [y_n(x)]^2 \, dx = \int_a^b f(x) y_n(x) dx \quad (n = 1, 2, \cdots). \tag{3.108}$$

In those cases when only a finite number of characteristic functions exist,

the functions generated by the operation

$$\int_a^b K(x,\xi)\Phi(\xi)d\xi$$

form a very restricted class. For example, if $K(x,\xi) = \sin(x+\xi)$ and $(a,b) = (0,2\pi)$, there follows

$$\begin{aligned}\int_0^{2\pi} K(x,\xi)\Phi(\xi)d\xi &= \int_0^{2\pi} (\sin x\cos\xi + \cos x\sin\xi)\Phi(\xi)d\xi \\ &= \left[\int_0^{2\pi} \Phi(\xi)\cos\xi d\xi\right]\sin x + \left[\int_0^{2\pi} \Phi(\xi)\sin\xi d\xi\right]\cos x,\end{aligned} \tag{3.109}$$

and hence this operation can generate only functions of the form

$$f(x) = C_1\sin x + C_2\cos x \tag{3.110}$$

regardless of the form of Φ. The characteristic functions of the associated homogeneous Fredholm integral equation

$$y(x) = \lambda\int_0^{2\pi} \sin(x+\xi)y(\xi)d\xi \tag{3.111}$$

are readily found, by the methods of the preceding section, to be arbitrary multiples of the functions $y_1(x) = \sin x + \cos x$ and $y_2(x) = \sin x - \cos x$, corresponding respectively to $\lambda_1 = 1/\pi$ and $\lambda_2 = -1/\pi$. It is obvious that any function of the form (3.110), generated by $\int_0^{2\pi}\sin(x+\xi)\Phi(\xi)d\xi$, can indeed be expressed as a linear combination of $y_1(x)$ and $y_2(x)$.

Even though the number of relevant independent characteristic functions be infinite, it is not necessarily true that any continuous function

$f(x)$ defined over (a, b) can be represented over that interval by a series of these functions; that is, the set of characteristic functions, even though infinite in number, may not comprise a *complete* set.

Suppose now that we have (in some manner) obtained all members of the set of normalized characteristic functions $y_n(x)$, each corresponding to a characteristic number λ_n of the homogeneous equation

$$y(x) = \lambda \int_a^b K(x, \xi) y(\xi) d\xi \tag{3.112}$$

where $K(x, \xi)$ is continuous, real, and symmetric. We next show that the knowledge of these functions and constants permits a simple determination of the continuous solution of the corresponding *nonhomogeneous* Fredholm equation of the second kind,

$$y(x) = F(x) + \lambda \int_a^b K(x, \xi) y(\xi) d\xi, \tag{3.113}$$

where $F(x)$ is a given continuous real function, when a continuous solution exists.

We suppose that the characteristic numbers have been ordered (say, with respect to magnitude), that characteristic numbers corresponding to k independent characteristic functions have been counted k times, and that such subsets of independent characteristic functions (corresponding to multiple characteristic numbers) have been orthogonalized.

In order to simplify the relations which follow, we suppose also that the arbitrary multiplicative constant associated with each characteristic function y_n is so chosen that the function is normalized over the relevant

interval (a, b). Thus we write

$$\phi_n = C_n y_n, \tag{3.114}$$

where the normalizing factor C_n is given by

$$C_n = \frac{1}{\sqrt{\int_a^b [y_n(x)]^2 \, dx}}, \tag{3.115}$$

so that there follows

$$\int_a^b [\phi_n(x)]^2 \, dx = 1. \tag{3.116}$$

The expansion (3.107) of a function $f(x)$ in a series of the normalized characteristic functions then takes the simpler form

$$f(x) = \sum_n a_n \phi_n(x) \quad \text{where} \quad a_n = \int_a^b f(x)\phi_n(x)dx. \tag{3.117}$$

Since the series (3.107) and (3.117) are identical, it follows that

$$a_n \phi_n = A_n y_n, \tag{3.118}$$

$$A_n = a_n C_n. \tag{3.119}$$

These relations permit transition from the expressions to be obtained to corresponding expressions involving normalized characteristic functions.

If the equation

$$y(x) = F(x) + \lambda \int_a^b K(x, \xi) y(\xi) d\xi \tag{3.120}$$

possesses a continuous solution $y(x)$, then the function $y(x) - F(x)$ is generated by the operation $\int_a^b K(x, \xi)[\lambda y(\xi)]d\xi$, and hence it can be rep-

resented by a series (or linear combination) of the normalized characteristic functions $\phi_n(x)(n = 1, 2, \cdots)$, of the form

$$y(x) - F(x) = \sum_n a_n \phi_n(x) \quad (a \leq x \leq b) \tag{3.121}$$

where the coefficients a_n are given, by virtue of (3.117), by

$$a_n = \int_a^b [y(x) - F(x)]\phi_n(x)dx. \tag{3.122}$$

With the convenient abbreviations

$$c_n = \int_a^b y(x)\phi_n(x)dx, \tag{3.123}$$

$$f_n = \int_a^b F(x)\phi_n(x)dx, \tag{3.124}$$

this relation takes the form

$$a_n = c_n - f_n. \tag{3.125}$$

In order to obtain a relation which permits the elimination of the unknown integral $c_n = \int_a^b y\phi_n dx$ from (3.125), we multiply both members of (3.120) by $\phi_n(x)$ and integrate the results over (a, b), so that there follows

$$c_n = f_n + \lambda \int_a^b \phi_n(x) \left[\int_a^b K(x, \xi)y(\xi)d\xi \right] dx, \tag{3.126}$$

with the notation of (3.123,3.124). If the order of integration is reversed in the coefficient of λ, and use is made of the assumed *symmetry* in

$K(x,\xi)$, that coefficient becomes

$$\int_a^b y(\xi)\left[\int_a^b K(\xi,x)\phi_n(x)dx\right]d\xi = \frac{1}{\lambda_n}\int_a^b y(\xi)\phi_n(\xi)d\xi = \frac{c_n}{\lambda_n},$$

and hence (3.126) is equivalent to the relation

$$c_n = f_n + \frac{\lambda}{\lambda_n}c_n. \tag{3.127}$$

The elimination of c_n between (3.125) and (3.127) then gives

$$a_n = \frac{\lambda}{\lambda_n - \lambda}f_n \tag{3.128}$$

if $\lambda \neq \lambda_n$. Hence the required solution (3.121) then takes the form

$$y(x) = F(x) + \lambda\sum_n \frac{f_n}{\lambda_n - \lambda}\phi_n(x) \quad (\lambda \neq \lambda_n) \tag{3.129}$$

where, as previously defined,

$$f_n = \int_a^b F(x)\phi_n(x)dx \quad (n = 1, 2, \cdots). \tag{3.130}$$

We may notice that the constants f_n would be the coefficients in the expansion $F(x) = \Sigma f_n\phi_n(x)$ if $F(x)$ were representable by such an expansion. It is of some importance to notice that in the preceding derivation it was not necessary to assume the validity of this representation.

The expansion (3.129) exists uniquely if and only if λ does not take on a characteristic value. Because of the presence of the term $f_k\phi_k(x)/(\lambda_k - \lambda)$, we see that if $\lambda = \lambda_k$, where λ_k is the k th characteristic number, the solution (3.129) is *nonexistent* unless also $f_k = 0$, that is, unless $F(x)$ is *orthogonal* to the corresponding characteristic func-

tion or functions. But if $\lambda = \lambda_k$ and $f_k = 0$ equation (3.127) reduces to the trivial identity when $n = k$, and hence imposes no restriction on c_k. From (3.125) it then follows that the coefficient of $\phi_k(x)$ in (3.129), which formally assumes the form $0/0$, is truly arbitrary, so that in this case (3.120) possesses infinitely many solutions, differing from each other by arbitrary multiples of $\phi_k(x)$. If λ assumes a characteristic value and $F(x)$ is not orthogonal to the corresponding characteristic function or functions, no solution exists.

By virtue of (3.118,3.119), the normalization of the characteristic functions is unnecessary, in the sense that (3.129) can be replaced by the expression

$$y(x) = F(x) + \lambda \sum_n \frac{F_n}{\lambda_n - \lambda} y_n(x) \quad (\lambda \neq \lambda_n)\,, \tag{3.131}$$

where

$$F_n \int_a^b [y_n(x)]^2\, dx = \int_a^b F(x) y_n(x) dx \quad (n = 1, 2, \cdots). \tag{3.132}$$

We consider next the Fredholm equation of the first kind,

$$F(x) = \int_a^b K(x, \xi) y(\xi) d\xi, \tag{3.133}$$

with a continuous, real, and symmetric kernel, where F is a prescribed continuous function and y is to be determined. It follows from the basic expansion theorem that (3.133) has no continuous solution unless $F(x)$ can be expressed as a linear combination of the characteristic functions corresponding to the associated homogeneous equation of the second

kind,

$$y(x) = \lambda \int_a^b K(x,\xi) y(\xi) d\xi. \tag{3.134}$$

For example, in the special case for which $K(x,\xi) = \sin(x+\xi)$ and $(a,b) = (0, 2\pi)$, equation (3.133) becomes

$$F(x) = \int_0^{2\pi} y(\xi) \sin(x+\xi) d\xi \tag{3.135}$$

or

$$F(x) = \left[\int_0^{2\pi} y(\xi) \cos\xi d\xi\right] \sin x + \left[\int_0^{2\pi} y(\xi) \sin\xi d\xi\right] \cos x. \tag{3.136}$$

This relation can be satisfied only if $F(x)$ is prescribed as a linear combination of $\sin x$ and $\cos x$ or, equivalently, as a related linear combination of the characteristic functions y_1 and y_2 of the associated homogeneous equation (3.111),

$$y_1 = \sin x + \cos x, \quad y_2 = \sin x - \cos x \tag{3.137}$$

corresponding to $\lambda_1 = 1/\pi$ and $\lambda_2 = -1/\pi$. If $F(x)$ is prescribed as

$$F(x) = A \sin x + B \cos x, \tag{3.138}$$

then (3.136) is satisfied by any function y for which

$$\int_0^{2\pi} y(\xi) \cos\xi d\xi = A, \quad \int_0^{2\pi} y(\xi) \sin\xi d\xi = B. \tag{3.139}$$

One such function is clearly

$$y(x) = \frac{1}{\pi}(A \cos x + B \sin x). \tag{3.140}$$

However, if we add to (3.140) any function which is orthogonal to both $\sin x$ and $\cos x$, and hence to the characteristic functions y_1 and y_2, over $(0, 2\pi)$, the conditions (3.139) will still be satisfied, so that the solution is by no means unique. Unless $F(x)$ is prescribed in the form (3.138), no solution exists.

Suppose now that (3.133) does possess a continuous solution. Then $F(x)$ is generated from $y(x)$ by the operation $\int_a^b K(x,\xi)y(\xi)d\xi$, and hence it can
be expanded in a series

$$F(x) = \sum_n f_n \phi_n(x) \quad (a \le x \le b), \tag{3.141}$$

where

$$f_n = \int_a^b F(x)\phi_n(x)dx \tag{3.142}$$

and where ϕ_n is the nth characteristic function of (3.134). The series may be finite or infinite. But since ϕ_n satisfies the equation

$$\phi_n(x) = \lambda_n \int_a^b K(x,\xi)\phi_n(\xi)d\xi, \tag{3.143}$$

and since (3.133) must be satisfied, we may replace $F(x)$ and $\phi_n(x)$ by the right-hand members of (3.133) and (3.143), so that (3.141) takes the form

$$\int_a^b K(x,\xi)y(\xi)d\xi = \sum_n \lambda_n f_n \int_a^b K(x,\xi)\phi_n(\xi)d\xi$$

or

$$\int_a^b K(x,\xi)\left[y(\xi) - \sum_n \lambda_n f_n \phi_n(\xi)\right] d\xi = 0. \tag{3.144}$$

This condition is satisfied if and only if $y(x)$ is of the form

$$y(x) = \sum_n \lambda_n f_n \phi_n(x) + \Phi(x), \tag{3.145}$$

where $\Phi(x)$ is a solution of the equation

$$\int_a^b K(x,\xi)\Phi(\xi)d\xi = 0. \tag{3.146}$$

We conclude that if (3.133) possesses a continuous solution, then that solution must be of the form (3.145), where Φ is any continuous function satisfying (3.146). From the homogeneity of (3.146) it is clear that either (3.146) is satisfied only by the trivial function $\Phi(x) \equiv 0$ or it possesses infinitely many solutions.

If we multiply both members of (3.146) by $\phi_n(x)$, integrate the results over (a,b) and make use of the assumed symmetry in K, we obtain the condition

$$\begin{aligned}\int_a^b \phi_n(x)\left[\int_a^b K(x,\xi)\Phi(\xi)d\xi\right]dx &= \int_a^b \Phi(\xi)\left[\int_a^b K(\xi,x)\phi_n(x)dx\right]d\xi \\ &= \frac{1}{\lambda_n}\int_a^b \Phi(\xi)\phi_n(\xi)d\xi = 0.\end{aligned} \tag{3.147}$$

Hence it follows that if (3.146) possesses a nontrivial solution, then that solution must be orthogonal to all the characteristic functions ϕ_n. If this set of functions is finite, then infinitely many linearly independent functions satisfying this condition exist. If the functions ϕ_n comprise an infinite complete set over (a,b), then no continuous nontrivial function can be simultaneously orthogonal to all functions of the set, so that in this case the function Φ in (3.145) must be identically zero.

In the preceding developments we made use of a known expansion theorem to show that if a Fredholm equation with a continuous, real, and symmetric kernel possesses a continuous solution, then that solution must be of a certain form. In particular, if the nonhomogeneous equation (3.113), of the second kind, possesses a continuous solution, then that solution is unique unless λ assumes a characteristic value, and is given by (3.129). If the equation (3.133), of the first kind, possesses a continuous solution, then it is given by (3.145) and it is or is not uniquely defined, according as (3.146) does not or does possess nontrivial solutions.

In the case of equation (3.113), it is known that a continuous solution does in fact always exist when $F(x)$ and $K(x, \xi)$ are continuous and λ is noncharacteristic, so that the validity of (3.129) then is completely established. However, as has been seen, no such assurance exists in the case of a Fredholm equation of the first kind.

In physically motivated problems, questions concerning existence and uniqueness of a solution usually can be resolved by physical considerations. Thus, for example, if the kernel in (3.146) is the Green's function

$$G(x,\xi) = \begin{cases} \delta \frac{x}{\zeta} & \text{when} \quad x < \zeta \\ \delta \frac{l-x}{l-\zeta} & \text{when} \quad x > \zeta \end{cases} \tag{3.148}$$

for a loaded string, a nontrivial solution of (3.146) clearly cannot exist, since it would represent a static loading which leads to no deflection at any point of the string. From the mathematical point of view, however, such questions are of considerable interest. Certain known results are included in the following sections.

It should be remarked that before the theory of this section can be applied it is necessary to determine the characteristic numbers and

functions of the homogeneous equation. Except in special cases this determination must depend upon numerical (or graphical) procedures.

Exercises

1. (a) Show that the kernel $K(x,\xi) = 1 + 3x\xi$ has a double characterestic number associated with $(-1, 1)$ with two independent characterestic functions.

   (b) Show that the kernel $K(x,\xi) = 1 + \xi + 3x\xi$ has a double characterestic number associated with $(-1, 1)$ with only one characterestic functions.

   (c) Show that the kernel $K(x,\xi) = \sin x \cos\xi$ has no characterestic numbers associated with $(0, 2\pi)$.

2. When λ is nearly equal to a characterestic number λ_m to which there corresponds only one characterestic function, show that the solution of the equation

$$y(x) = F(x) + \lambda \int_a^b K(x,\xi) y(\xi) d\xi,$$

   where $K(x,\xi)$ is real and symmetric, is given approximately by

$$y(x) \approx F(x) + \frac{\lambda}{\lambda_m - \lambda} \left(\int_a^b F_{\phi_m} dx \right) \phi_m(x),$$

   where $\phi_m(x)$ is the corresponding normalized characterestic function.

3.7 Iterative methods for solving equations of the second kind

In certain cases integral equations of the second kind can be solved by a method of successive approximations. In this section we describe the method and investigate its validity.

Suppose that in a Fredholm equation of the second kind,

$$y(x) = F(x) + \lambda \int_a^b K(x,\xi) y(\xi) d\xi \tag{3.149}$$

where F and K are continuous, we replace y under the integral sign by an initial approximation $y^{(0)}$. Then (3.149) determines an approximation $y^{(1)}$ in the form

$$y^{(1)}(x) = F(x) + \lambda \int_a^b K(x,\xi) y^{(0)}(\xi) d\xi \tag{3.150}$$

By substituting this approximation into the right-hand member of (3.149), we may then obtain the next approximation $y^{(2)}$, and continue the process in such a way that successive approximations are determined by the formula

$$y^{(n)}(x) = F(x) + \lambda \int_a^b K(x,\xi) y^{(n-1)}(\xi) d\xi. \tag{3.151}$$

The same method is clearly applicable also when the upper limit b is replaced by the current variable x, so that the equation is of the *Volterra type*. It remains to determine under what conditions the successive approximations actually tend toward a continuous solution of (3.149).

In order to examine this procedure more closely, we write out explicitly the results of the indicated substitutions. Thus we first obtain

the result of replacing y in the right-hand member of (3.149) by $y^{(1)}$, as given by (3.150). In this substitution, we must replace the current variable x in (3.150) by the dummy variable ξ appearing in (3.149). To avoid ambiguity, we must then replace ξ in (3.150) by another dummy variable, say ξ_1, so that (3.150) becomes

$$y^{(1)}(\xi) = F(\xi) + \lambda \int_a^b K(\xi, \xi_1)\, y^{(0)}(\xi_1)\, d\xi_1.$$

The result of the substitution then takes the form

$$y^{(2)}(x) = F(x) + \lambda \int_a^b K(x, \xi) \left[F(\xi) + \lambda \int_a^b K(\xi, \xi_1)\, y^{(0)}(\xi_1)\, d\xi_1 \right] d\xi$$

or

$$\begin{aligned} y^{(2)}(x) = F(x) + \lambda \int_a^b K(x, \xi) F(\xi) d\xi + \\ \lambda^2 \int_a^b K(x, \xi) \int_a^b K(\xi, \xi_1)\, y^{(0)}(\xi_1)\, d\xi_1 d\xi. \end{aligned} \tag{3.152}$$

If we now replace x by ξ, ξ by ξ_1, and ξ_1 by ξ_2, in (3.152), and substitute once more in the right-hand member of (3.149), there follows

$$\begin{aligned} y^{(3)}(x) =& F(x) + \lambda \int_a^b K(x, \xi) F(\xi) d\xi \\ &+ \lambda^2 \int_a^b K(x, \xi) \int_a^b K(\xi, \xi_1)\, F(\xi_1)\, d\xi_1 d\xi + \\ \lambda^3 \int_a^b K(x, \xi) \int_a^b K(\xi, \xi_1) \int_a^b K(\xi_1, \xi_2)\, y^{(0)}(\xi_2)\, d\xi_2 d\xi_1 d\xi. \end{aligned} \tag{3.153}$$

The analysis is abbreviated considerably if we introduce an *integral*

operator $\mathscr{K}$, defined by the equation

$$\mathscr{K} f(x) \equiv \int_a^b K(x,\xi) f(\xi) d\xi. \tag{3.154}$$

The integral equation (3.149) then takes the symbolic form

$$y(x) = F(x) + \lambda \mathscr{K} y(x), \tag{3.155}$$

while (3.151) becomes

$$y^{(n)}(x) = F(x) + \lambda \mathscr{K} y^{(n-1)}(x). \tag{3.156}$$

Further, equations (3.150), (3.152), and (3.153) take the forms

$$\left.\begin{aligned} y^{(1)}(x) &= F(x) + \lambda \mathscr{K} y^{(0)}(x) \\ y^{(2)}(x) &= F(x) + \lambda \mathscr{K} F(x) + \lambda^2 \mathscr{K}^2 y^{(0)}(x), \\ y^{(3)}(x) &= F(x) + \lambda \mathscr{K} F(x) + \lambda^2 \mathscr{K}^2 F(x) + \lambda^3 \mathscr{K}^3 y^{(0)}(x) \end{aligned}\right\}. \tag{3.157}$$

More generally, after the nth substitution we have

$$y^{(n)}(x) = F(x) + \lambda \mathscr{K} F(x) + \lambda^2 \mathscr{K}^2 F(x) + \lambda^3 \mathscr{K}^3 F(x) + \cdots + \lambda^{n-1} \mathscr{K}^{n-1} F(x) + R_n(x) \tag{3.158}$$

where $R_n(x)$ is defined by

$$R_n(x) = \lambda^n \mathscr{K}^n y^{(0)}(x). \tag{3.159}$$

Hence, as $n \to \infty$, we are led to the possibility that the desired solution

of (3.149) can be expressed as the infinite series

$$y(x) = F(x) + \sum_{n=1}^{\infty} \lambda^n \mathscr{K}^n F(x), \tag{3.160}$$

It remains to determine conditions under which the expression $R_n(x)$ tends to zero and under which the formal series (3.160) actually converges, and represents a continuous solution of (3.149).

Since $K(x, \xi)$ and $F(x)$ are assumed to be continuous for all values of x and ξ in (a, b), their magnitudes certainly are bounded in (a, b), so that there exist positive constants M and m such that

$$|K(x, \xi)| \leq M, \quad |F(x)| \leq m \tag{3.161}$$

in (a, b). We suppose also that the magnitude of the initial approximation $y^{(0)}(x)$ is bounded in (a, b), so that

$$\left| y^{(0)}(x) \right| \leq C \tag{3.162}$$

in that interval.

With the understanding that $b > a$, there then follows

$$\left| \mathscr{K} y^{(0)}(x) \right| = \left| \int_a^b K(x, \xi) y^{(0)}(\xi) d\xi \right| \leq \int_a^b MC d\xi = M(b - a)C.$$

More generally, we find by iteration that

$$\left| \mathscr{K}^n y^{(0)}(x) \right| \leq M^n (b - a)^n C \tag{3.163}$$

and, similarly,

$$|\mathscr{K}^n F(x)| \leq M^n (b - a)^n m. \tag{3.164}$$

Hence, according to (3.159) and (3.163), there follows

$$|R_n(x)| \leq |\lambda|^n M^n (b-a)^n C, \tag{3.165}$$

and we may deduce that $R_n(x)$ tends to zero with increasing n if

$$|\lambda| < \frac{1}{M(b-a)}. \tag{3.166}$$

Further, from (3.164) it follows that the series

$$|F(x)| + \sum_{n=1}^{\infty} |\lambda|^n |\mathscr{K}^n F(x)|$$

is dominated by the constant series

$$m\left[1 + \sum_{n}^{\infty} |\lambda|^n M^n (b-a)^n\right].$$

Since this geometric series converges when λ satisfies the inequality (3.166), we may deduce that the series (3.160) converges absolutely and uniformly in (a, b) when (3.166) is satisfied.

It is then easily seen, by direct substitution and term-by-term integration, that the series (3.160) actually satisfies the integral equation (3.149) and hence represents the continuous solution of (3.149) when (3.166) is satisfied and K is continuous.

The series (3.160) is a power series in λ. If we recall that the solution to (3.149) generally fails to exist when λ takes on a characteristic value, we are led to expect that the series solution (3.160) will cease to converge at least as soon as $|\lambda|$ becomes equal to the absolute value of the smallest characteristic number λ_1. It can be shown that this is the case and,

indeed, that the series (3.160) converges when $|\lambda| < |\lambda_1|$, and only then.

Since the condition (3.166) may be conservative, in the sense that the series (3.160) may perhaps converge even though (3.166) is not satisfied, we thus obtain as a by-product of this derivation the useful relation

$$|\lambda_1| \geq \frac{1}{M(b-a)}, \tag{3.167}$$

which gives a lower bound for the magnitude of the smallest characteristic number λ_1. When K is real and symmetric, a somewhat more involved analysis leads to the inequality

$$|\lambda_1| \geq \frac{1}{\sqrt{\int_a^b \int_a^b [K(x,\xi)]^2 dx d\xi}} \tag{3.168}$$

which gives a sharper lower bound in that case.

It is of some interest to notice that the relation (3.160) can also be obtained by writing (3.155) in the operational form

$$(\mathscr{I} - \lambda \mathscr{K}) y(x) = F(x), \tag{3.169}$$

where $\mathscr{I}$ is the identity operator, deducing that

$$y(x) = (\mathscr{I} - \lambda \mathscr{K})^{-1} F(x) \tag{3.170}$$

where $(\mathscr{I} - \lambda \mathscr{K})^{-1}$ represents an operator inverse to $\mathscr{I} - \lambda \mathscr{K}$, and formally applying the binomial theorem to deduce that

$$(\mathscr{I} - \lambda \mathscr{K})^{-1} = \mathscr{I} + \lambda \mathscr{K} + \lambda^2 \mathscr{K}^2 + \cdots, \tag{3.171}$$

Only when $|\lambda| < |\lambda_1|$ is this formal result valid.

In the case of the *Volterra* equation,

$$y(x) = F(x) + \lambda \int_a^x K(x,\xi) y(\xi) d\xi, \tag{3.172}$$

with a variable upper limit, we define $\mathscr{K}_x$ as the integral operator such that

$$\mathscr{K}_x f(x) \equiv \int_a^x K(x,\xi) f(\xi) d\xi, \tag{3.173}$$

and a procedure completely analogous to that used above leads to the formal solution

$$y(x) = F(x) + \sum_{n=1}^{\infty} \lambda^n \mathscr{K}_x^n F(x), \tag{3.174}$$

if the expression

$$R_n(x) = \lambda^n \mathscr{K}_x^n y^{(0)}(x) \tag{3.175}$$

tends to zero as $n \to \infty$.

In this case, if we consider any interval (a, b), where b is any number larger than a, and again assume the bounds (3.161) and (3.162) over (a, b), we have

$$\left| \mathscr{K}_x y^{(0)}(x) \right| = \left| \int_a^x K(x,\xi) y^{(0)}(\xi) d\xi \right| \leq MC \int_a^x d\xi = MC(x-a),$$

when $a \leq x \leq b$, and hence also

$$\begin{aligned} \left| \mathscr{K}_x^2 y^{(0)}(x) \right| &\leq \int_a^x |K(x,\xi)| MC(\xi - a) d\xi \\ &\leq M^2 C \int_a^x (\xi - a) d\xi = \frac{M^2 (x-a)^2}{2 \cdot 1} C. \end{aligned}$$

Inductive reasoning then leads to the result

$$\begin{aligned} |R_n(x)| = \left|\lambda^n \mathscr{N}_x^n y^{(0)}(x)\right| &\leq |\lambda|^n \frac{M^n (x-a)^n}{n!} C \\ &\leq |\lambda|^n \frac{M^n (b-a)^n}{n!} C \end{aligned} \tag{3.176}$$

for $a \leq x \leq b$. In a similar way, it is found that

$$|\lambda^n \mathscr{K}_x^n F(x)| \leq |\lambda|^n \frac{M^n (b-a)^n}{n!} m \tag{3.177}$$

for $a \leq x \leq b$. But the last member of (3.176) tends to zero as $n \to \infty$, and also the series whose nth term is given by the right-hand member of (3.177) converges, for any finite value of λ. Thus the method of successive substitutions converges to the series (3.174) and that series converges absolutely and uniformly, for any finite value of λ, in any interval (a, b) for which $b > a$. A similar result follows for any $b < a$. It then follows, by direct substitution, that the series (3.174) converges to the unique continuous solution of the Volterra equation (3.172) for all values of λ, in any interval (a, b) in which $F(x)$ and $K(x, \xi)$ are continuous.

We may notice that, as was to be expected, the final solution in each case is independent of the initial approximation $y^{(0)}(x)$. In practice, it is often desirable to merely evaluate successive terms in the series (3.160) or (3.174) by iteration, as is illustrated by an example which follows, rather than to actually pursue the method of successive substitutions which motivated (3.160) and (3.174). The latter method reduces to the former when the initial approximation

$$y^{(0)}(x) = F(x)$$

is used. However, when advance information as to the nature of the

required solution is available, a better choice of $y^{(0)}(x)$ may suggest itself.

As a very simple illustration of these results, we consider the Fredholm equation

$$y(x) = 1 + \lambda \int_0^1 (1 - 3x\xi) y(\xi) d\xi, \tag{3.178}$$

the solution of which is in the form

$$y(x) = \frac{4 + 2\lambda(2 - 3x)}{4 - \lambda^2} \quad (\lambda \neq \pm 2). \tag{3.179}$$

The operation $\mathscr{K} f(x)$ is then of the form

$$\mathscr{K} f(x) = \int_0^1 (1 - 3x\xi) f(\xi) d\xi.$$

To obtain the series solution (3.160), we make the calculations

$$\begin{aligned}
\mathscr{K} F &= \int_0^1 (1 - 3x\xi) d\xi = 1 - \frac{3}{2}x, \\
\mathscr{K}^2 F &= \int_0^1 (1 - 3x\xi) \left(1 - \frac{3}{2}\xi\right) d\xi = \frac{1}{4}, \\
\mathscr{K}^3 F &= \int_0^1 (1 - 3x\xi) \frac{1}{4} d\xi = \frac{1}{4}\left(1 - \frac{3}{2}x\right),
\end{aligned}$$

and so forth. From the form of these results the form of the general result is obvious, and (3.160) becomes

$$\begin{aligned}
y(x) = 1 + \lambda \left(1 - \frac{3}{2}x\right) + \frac{\lambda^2}{4} + \frac{\lambda^3}{4}\left(1 - \frac{3}{2}x\right) \\
+ \frac{\lambda^4}{16} + \frac{\lambda^5}{16}\left(1 - \frac{3}{2}x\right) + \cdots
\end{aligned} \tag{3.180}$$

This result can be expressed in the form

$$y(x) = \left(1 + \frac{\lambda^2}{4} + \frac{\lambda^4}{16} + \cdots\right)\left[1 + \lambda\left(1 - \frac{3}{2}x\right)\right]. \tag{3.181}$$

The power series in (3.181) is a geometric series, convergent when $|\lambda| < 2$, with the sum $1/\left(1 - \lambda^2/4\right)$. Hence (3.180,3.181) is the power series expansion, valid when and only when $|\lambda| < 2$, of the solution

$$y(x) = \frac{1 + \lambda\left(1 - \frac{3}{2}x\right)}{1 - \frac{1}{4}\lambda^2}, \tag{3.182}$$

which is identical with (3.179), and which itself is valid for all values of λ except $\lambda = \pm 2$.

We may notice that consequently the method of successive substitutions would not converge, for example, if it were applied to the integral equation

$$y(x) = 1 + 4\int_0^1 (1 - 3x\xi)y(\xi)d\xi,$$

while it would converge if applied to the same equation with the factor 4 replaced by any number smaller than two in absolute value.

Obviously the method of Section 3.4 is generally to be preferred when the kernel $K(x, \xi)$ is separable, as in the preceding case. It is important to notice that the methods of Section 3.4 express the solution as the ratio of a function of x and λ to a polynomial $\Delta(\lambda)$, valid for all values of λ except the characteristic values for which $\Delta(\lambda) = 0$. The method of successive substitutions, in such cases, leads to the power series expansion of this ratio, valid only when $|\lambda|$ is smaller than the magnitude of the smallest characteristic number.

In the more important cases when the kernel is not separable, there

exists a method, due to Fredholm, which generalizes the procedure of Section 3.4. This method is discussed in Section 3.9.

In the following section the preceding developments a re considered from a slightly different viewpoint.

Exercises

1. Consider the integral equation

$$y(x) = \lambda \int_0^1 x\xi y(\xi) d\xi + 1.$$

   (a) Make use of equation (3.168) to show in advance that the iterative procedure of section 3.7 will converge when $|\lambda| < 3$.

   (b) Show that the iterative procedure leads formally to the expression

$$y(x) = 1 + x\left(\frac{\lambda}{2} + \frac{\lambda^2}{6} + \frac{\lambda^3}{18} + \cdots\right)$$

   (c) Use the method of section 3.4 to obtain the exact solution of the problem in the form

$$y(x) = 1 + \frac{3\lambda x}{2(3-\lambda)} \quad (\lambda \neq 3).$$

2. Deal with the integral equation

$$y(x) = \lambda \int_0^1 (x + \xi) y(\xi) d\xi + 1.$$

   as in the above problem 1. Show also that the estimate afforded by (3.168) is slightly conservative in this case.

3. Deal with the integral equation

$$y(x) = \sin x + \lambda \int_0^{2\pi} \cos(x+\xi) y(\xi) d\xi.$$

as in the above problem 1.

3.8 The Neumann series

With the notation of equation (3.154),

$$\mathscr{K} f(x) = \int_a^b K(x,\xi) f(\xi) d\xi, \tag{3.183}$$

there follows also

$$\begin{aligned} \mathscr{K}^2 f(x) &= \int_a^b K(x,\xi_1)\, \mathscr{N} f(\xi_1)\, d\xi_1 \\ &= \int_a^b K(x,\xi_1) \left[\int_a^b K(\xi_1,\xi) f(\xi) d\xi \right] d\xi_1 \\ &= \int_a^b \left[\int_a^b K(x,\xi_1) K(\xi_1,\xi)\, d\xi_1 \right] f(\xi) d\xi \end{aligned} \tag{3.184}$$

If we define the *iterated kernel* $K_2(x,\xi)$ by the relation

$$K_2(x,\xi) = \int_a^b K(x,\xi_1) K(\xi_1,\xi)\, d\xi_1, \tag{3.185}$$

equation (3.184) takes the form

$$\mathscr{K}^2 f(x) = \int_a^b K_2(x,\xi) f(\xi) d\xi. \tag{3.186}$$

By repeating this process, it is easily seen that one can write

$$\mathscr{K}^n f(x) = \int_a^b K_n(x,\xi) f(\xi) d\xi, \tag{3.187}$$

where $K_n(x,\xi)$ is the nth iterated kernel, defined by the recurrence formula

$$K_n(x,\xi) = \int_a^b K\,(x,\xi_1)\,K_{n-1}\,(\xi_1,\xi)\,d\xi_1, \tag{3.188}$$

for $n = 2, 3, 4, \cdots$, and where we write

$$K_1(x,\xi) \equiv K(x,\xi). \tag{3.189}$$

It is not difficult to establish the consequent validity of the relation

$$K_{p+q}(x,\xi) = \int_a^b K_p\,(x,\xi_1)\,K_q\,(\xi_1,\xi)\,d\xi_1, \tag{3.190}$$

for any positive integers p and q. Further, if $K(x,\xi)$ is bounded in (a,b), in such a way that

$$|K(x,\xi)| \leq M \tag{3.191}$$

in (a,b), then it follows easily that also

$$|K_n(x.\xi)| \leq M^n (b-a)^{n-1} \tag{3.192}$$

for values of x and ξ in (a,b).

With the notation of (3.187), the series (3.160), representing the solution of the equation

$$y(x) = F(x) + \lambda \int_a^b K(x,\xi) y(\xi) d\xi \tag{3.193}$$

for sufficiently small values of $|\lambda|$, takes the form

$$\begin{aligned} y(x) &= F(x) + \sum_{n=1}^{r} \lambda^n \int_a^b K_n(x,\xi) F(\xi) d\xi \\ &= F(x) + \lambda \int_a^b \left[\sum_{n=0}^{k} \lambda^n K_{n+1}(x,\xi) \right] F(\xi) d\xi, \end{aligned} \tag{3.194}$$

assuming the legitimacy of interchange of summation and integration. If we introduce the abbreviation

$$\begin{aligned} \Gamma(x,\xi;\lambda) &= \sum_{n=0}^{\infty} \lambda^n K_{n+1}(x,\xi) \\ &= K(x,\xi) + \lambda K_2(x,\xi) + \lambda^2 K_3(x,\xi) + \cdots, \end{aligned} \tag{3.195}$$

equation (3.194) takes the form

$$y(x) = F(x) + \lambda \int_a^b \Gamma(x,\xi;\lambda) F(\xi) d\xi. \tag{3.196}$$

The function $\Gamma(x,\xi;\lambda)$ is known as the reciprocal or resolvent kernel associated with the kernel $K(x,\xi)$ in the interval (a,b). Further, the series (3.195) [or, in some references, the series (3.194)] is known as the Neumann series. If use is made of (3.192), it is found that this series converges (absolutely and uniformly) when $|\lambda| < 1/M(b-a)$. A more precise analysis shows indeed that the series converges when $|\lambda| < |\lambda_1|$, where λ_1 is the smallest characteristic number, as was the case in the analogous expansions of the preceding section. In fact, equation (3.196) is merely an abbreviation for (3.194), which is equivalent to (3.160) by virtue of (3.187).

In practice, unless the solution of (3.193) is required for several choices of $F(x)$, it may be more convenient to obtain leading terms of

the series (3.194) or (3.160) by the iterative methods of the preceding section, than actually to evaluate terms of (3.195) and insert the result into (3.196); the net results are, of course, the same. However, the Neumann series and the resolvent kernel are of importance in theoretical developments. An interesting result in this connection is obtained if we rewrite (3.195) in the form

$$\begin{aligned}\Gamma(x,\xi;\lambda) &= K(x,\xi) + \lambda \sum_{n-0}^{\infty} \lambda^n K_{n+2}(x,\xi) \\ &= K(x,\xi) + \lambda \sum_{n=0}^{\infty} \lambda^n \int_a^b K(x,\xi_1)\, K_{n+1}(\xi_1,\xi)\, d\xi_1\end{aligned}$$

and hence, again referring to (3.195), deduce the relation

$$\Gamma(x,\xi;\lambda) = K(x,\xi) + \lambda \int_a^b K(x,\xi_1)\, \Gamma(\xi_1,\xi;\lambda)\, d\xi_1,$$

or, with a change in notation,

$$\Gamma(x,t;\lambda) = K(x,t) + \lambda \int_a^b K(x,\xi)\Gamma(\xi,t;\lambda) d\xi. \tag{3.197}$$

Thus it follows that the resolvent kernel Γ, considered as a function of the two variables x and t and the parameter λ, is the solution of equation (3.193) when the prescribed function F is replaced by the kernel K, considered as a function of x and t.

In order to illustrate the actual determination of the resolvent kernel in a very simple case, we again consider equation (3.178). With

$$K(x,\xi) = 1 - 3x\xi,$$

there follows

$$K_2(x,\xi) = \int_0^1 (1-3x\xi_1)\,(1-3\xi_1\xi)\,d\xi_1 = 1 - \frac{3}{2}(x+\xi) + 3x\xi.$$

and, similarly,

$$K_3(x,\xi) = \int_0^1 K\,(x,\xi_1)\,K_2\,(\xi_1,\xi)\,d\xi_1 = \frac{1}{4}(1-3x\xi).$$

Since, in this special case, we therefore have $K_3 = K_1/4$, it follows easily that $K_n = K_{n-2}/4$ for $n \geq 3$, and hence we have

$$\begin{aligned}\Gamma &= K_1 + \lambda K_2 + \lambda^2 K_3 + \cdots \\ &= \left(1 + \frac{\lambda^2}{4} + \frac{\lambda^4}{16} + \cdots\right) K_1 + \lambda \left(1 + \frac{\lambda^2}{4} + \frac{\lambda^4}{16} + \cdots\right) K_2\end{aligned}$$

or

$$\Gamma(x,\xi;\lambda) = \frac{1}{1-\lambda^2/4}\left[(1+\lambda) - \frac{3}{2}\lambda(x+\xi) - 3(1-\hat{\lambda})x\xi\right] \quad (|\lambda| < 2). \tag{3.198}$$

The introduction of this function into (3.196), with $F(x) = 1$, leads again to the solution (3.182).

It is important to notice that the result obtained is correct for all values of λ except $\lambda = \pm 2$. That is, the resolvent kernel is correctly given by (3.198) for all such values of λ. However, the series involved in the equation preceding (3.198) converges only when $|\lambda| < 2$. It happens that we are able to sum that series explicitly in the present example, and that the resultant function correctly represents the resolvent kernel for all values of λ other than characteristic values.

Exercises

1. Determine the resolvent kernel associated with $K(x, \xi) = x\xi$ in the interval $(0, 1)$ in the form of a power series in λ, obtaining the first three terms.

2. Determine the resolvent kernel associated with $K(x, \xi) = x + \xi$ in the interval $(0, 1)$ in the form of a power series in λ, obtaining the first three terms.

3. Determine the resolvent kernel associated with $K(x, \xi) = \cos(x+\xi)$ in the interval $(0, 2\pi)$ in the form of a power series in λ.

4. Determine the iterated kernel $K_2(x, \xi)$ associated with $K(x, \xi) = |x - \xi|$ in the interval $(0, 1)$.

3.9 Fredholm theory

In the general case of a continuous kernel $K(x, \xi)$, which need not be real, it is possible to express the resolvent kernel $\Gamma(x, \xi; \lambda)$ as the ratio of two infinite series of powers of λ, in such a way that both series converge for all values of λ. The derivation of the basic equations, due originally to Fredholm, involves considerable algebraic manipulation and is not considered here.

If the resolvent kernel is expressed as the ratio

$$\Gamma(x, \xi; \lambda) = \frac{D(x, \xi; \lambda)}{\Delta(\lambda)}, \tag{3.199}$$

where

$$D(x,\xi;\lambda) = K(x,\xi) + \lambda D_1(x,\xi) + \lambda^2 D_2(x,\xi) + \cdots \tag{3.200}$$

and

$$\Delta(\lambda) = 1 + \lambda C_1 + \lambda^2 C_2 + \cdots, \tag{3.201}$$

it is found that the coefficients C_n and the functions $D_n(x,\xi)$ can be determined successively by the following sequence of calculations:

$$C_1 = \int_a^b K(x,x)dx, \;\; D_1(x,\xi) = C_1 K(x,\xi) + \int_a^b K\left(x,\xi_1\right) K\left(\xi_1,\xi\right) d\xi_1;$$
$$C_2 = \int_a^b D_1(x,x)dx, \;\; D_2(x,\xi) = C_2 K(x,\xi) + \int_a^b K\left(x,\xi_1\right) D_1\left(\xi_1,\xi\right) d\xi_1;$$
.....
$$C_n = \int_a^b D_{n-1}(x,x)dx, \qquad D_n(x,\xi) = C_n K(x,\xi) + \int_a^b K\left(x,\xi_1\right) D_{n-1}\left(\xi_1,\xi\right) d\xi_1. \tag{3.202}$$

The solution of the equation

$$y(x) = F(x) + \lambda \int_a^b K(x,\xi) y(\xi) d\xi \tag{3.203}$$

is then obtained by introducing (3.199) in to the equation,

$$y(x) = F(x) + \lambda \int_a^b \Gamma(x,\xi;\lambda) F(\xi) d\xi,$$

in the form

$$y(x) = F(x) + \lambda \frac{\int_a^b D(x, \xi; \lambda) F(\xi) d\xi}{\Delta(\lambda)} \tag{3.204}$$

In those cases when $K(x, \xi)$ is separable, this result is identical in form with the solution obtained by the methods of Section 3.4 The series (3.200) and (3.201) then each involve only a finite number of terms.

More generally, if the ratio of the two power series involved in 3.204 were expressed as a single power series in λ (by division or otherwise). However, the result of this operation would converge only for small values of $|\lambda|$ (when $|\lambda| < |\lambda_1|$), whereas the separate series expansions of the numerator and denominator in the last term of (3.204) each converge for all values of λ.

The denominator $\Delta(\lambda)$ vanishes only when λ takes on a characteristic value, in which case either no solution or infinitely many solutions of (3.203) exist, and (3.204) is no longer valid.

Despite the generality of the solution just described, the practical usefulness of the result is limited by the fact that the relevant calculations usually involve a prohibitive amount of labor unless $K(x, \xi)$ is separable (and hence the simpler methods of Section 3.4 are usually preferable). Nevertheless, the rigorous development of the underlying theory has led to valuable information concerning existence and uniqueness of solutions of (3.203). In the following paragraph we summarize certain known facts which generalize results already obtained in the special cases when the kernel is either separable or real and symmetric.

The equation

$$y(x) = F(x) + \lambda \int_a^b K(x,\xi) y(\xi) d\xi, \tag{3.205}$$

where (a, b) is a finite interval and where $F(x)$ and $K(x, \xi)$ are continuous in (a, b), possesses one and only one continuous solution for any fixed value of λ which is not a characteristic value. If λ_c is a characteristic number of multiplicity r, that is, if the associated homogeneous equation

$$y(x) = \lambda_c \int_a^b K(\xi, x) y(\xi) d\xi \tag{3.206}$$

possesses r linearly independent nontrivial solutions $\phi_1, \phi_2, \cdots, \phi_r$, then r is finite and the associated transposed homogeneous equation

$$z(x) = \lambda_c \int_a^b K(\xi, x) z(\xi) d\xi \tag{3.207}$$

also possesses r linearly independent nontrivial solutions $\psi_1, \psi_2, \cdots, \psi_r$. In this exceptional case, (3.205) possesses no solution unless $F(x)$ is orthogonal to each of the characteristic functions $\psi_1, \psi_2, \cdots, \psi_r$,

$$\int_a^b F(x)\psi_k(x)dx = 0 \quad (k = 1, 2, \cdots, r) \tag{3.208}$$

Finally, if $\lambda = \lambda_c$ and (3.208) is satisfied, then the solution of (3.205) is determinate only within an additive linear combination $c_1\phi_1 + c_2\phi_2 + \cdots + c_r\phi_r$, where the r constants c_n are arbitrary.

A Volterra integral equation, of the form

$$y(x) = F(x) + \lambda \int_a^x K(x,\xi) y(\xi) d\xi, \tag{3.209}$$

can be considered as a special form of a Fredholm equation, with a kernel given by the expressions

$$\bar{K}(x,\xi) = \begin{cases} 0 & x < \xi, \\ K(x,\xi) & x \geq \xi. \end{cases} \tag{3.210}$$

However, unless $K(x,x) = 0$, the modified kernel $\widetilde{K}(x,\xi)$ is discontinuous when $x = \xi$. The results of Section 3.7 show that if $F(x)$ and $K(x,\xi)$ are continuous the Volterra equation (3.209) possesses one and only one contimuous solution. In particular, when $F(x) \equiv 0$ the only possible continuous solution of (3.209) is then the trivial solution $y(x) \equiv 0$.

Exercises

1. Obtain the resolvent kernel associated with $K(x,\xi) = x\xi$ in $(0,1)$, by the method of Section 3.9. For the kernel $K(x,\xi) - x + \xi$ in $(0,1)$.

2. (a) Obtain the solution of the equation

$$y(x) = \lambda \int_a^b K(x,\xi) y(\xi) d\xi + F(x),$$

where

$$K(x,\xi) = u(x)v(\xi),$$

in the form

$$y(x) = F(x) + \lambda \frac{\int_a^b K(x,\xi)F(\xi)d\xi}{1 - \lambda \int_a^b K(x,x)dx},$$

by the methods of Section 3.4. [Write $y(x) = \lambda\, \mathrm{cu}(x) + F(x)$, where

$$c = \int_a^b v(\xi)y(\xi)d\xi,$$

determine c, and identify the result with the form given.]

(b) Obtain the same result by the methods of Section 3.9. [Notice that when $\int_a^b K(x,x)dx = 0$ there is no characteristic number, whereas otherwise the single characteristic number is $\lambda_1 = 1/\int_a^b K(x,x)dx$.]