1 Introduction and finite-difference formulae

The mathematical formulation of most problems in science involving rates of change with respect to two or more independent variables, usually representing time, length or angle, leads either to a partial differential equation or to a set of such equations. Special cases of the two dimensional second-order equation

$$a\frac{\partial^2\phi}{\partial x^2}+b\frac{\partial^2\phi}{\partial x\,\partial y}+c\frac{\partial^2\phi}{\partial y^2}+d\frac{\partial\phi}{\partial x}+e\frac{\partial\phi}{\partial y}+f\phi+g=0,$$

where a, b, c, d, e, f, and g may be functions of the independent variables x and y and of the dependent variable ϕ, occur more frequently than any other because they are often the mathematical form of one of the conservation principles of physics.

For reasons that are given in Chapter 4 this equation is said to be *elliptic* when $b^2-4ac<0$, *parabolic* when $b^2-4ac=0$, and *hyperbolic* when $b^2-4ac>0$.

Two-dimensional elliptic equations

These equations, of which the best known are Poisson's equation

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

and Laplace's equation

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

are generally associated with equilibrium or steady-state problems. For example, the velocity potential for the steady flow of incompressible non-viscous fluid satisfies Laplace's equation and is the mathematical way of expressing the idea that the rate at which such fluid enters any given region is equal to the rate at which it leaves it. Similarly, the electric potential V associated with a two-dimensional electron distribution of charge density ρ

satisfies Poisson's equation $\partial^2 V/\partial x^2 + \partial^2 V/\partial y^2 + \rho/\varepsilon = 0$, where ε is the dielectric constant. This is the partial differential equation form of the well-known theorem by Gauss which states that the total electric flux through any closed surface is equal to the total charge enclosed.

The *analytical* solution of a two-dimensional elliptic equation is a function of the space co-ordinates x and y which satisfies the partial differential equation *at every point* of the area S inside a plane *closed* curve C and satisfies certain conditions *at every point* on this boundary curve C (Fig. 1.1). The function ϕ, for instance, from which we can calculate the displacements and shear stresses within a long solid elastic cylinder in a state of torsion satisfies

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

at every point of a right cross-section, and has a constant value round the perimeter of the cross-section. Similarly, the steady motion of incompressible viscous fluid through a straight uniform tube can be found from a function that satisfies Laplace's equation at every point of the cross-section and equals $\frac{1}{2}(x^2 + y^2)$ at each point on the boundary.

The condition that the dependent variable must satisfy round the boundary curve C is termed the boundary condition.

To the present, only a limited number of special types of

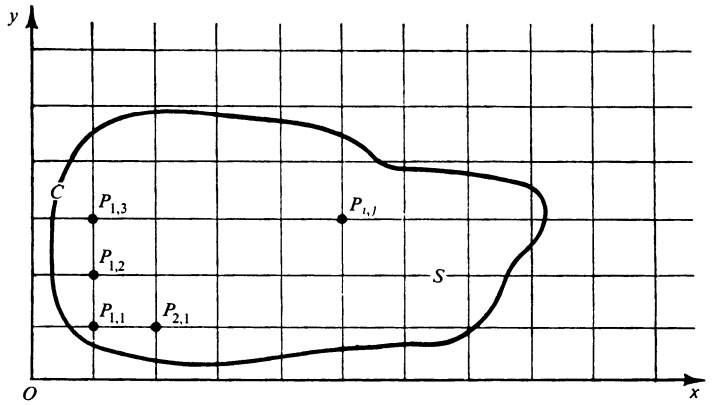

Fig. 1.1

elliptic equations have been solved analytically and the usefulness of these solutions is further restricted to problems involving shapes for which the boundary conditions can be satisfied. This not only eliminates all problems with boundary curves that are undefined in terms of equations, but also many for which the boundary conditions are too difficult to satisfy even though the equations for the boundary curves are known. In such cases approximation methods, whether analytical or numerical in character, are the only means of solution, apart from the use of analogue devices. Analytical approximation methods often provide extremely useful information concerning the character of the solution for critical values of the dependent variables but tend to be more difficult to apply than the numerical methods, and will not be discussed in this book. Of the numerical approximation methods available for solving differential equations those employing finite-differences or finite elements are more frequently used and more universally applicable than any other, although finite elements are not considered in this book. Before outlining these methods however, the reader should be aware of the manner in which the term 'approximation method' is used. Finite-difference methods are approximate in the sense that derivatives *at a point* are approximated by difference quotients *over a small interval*, i.e., $\partial\phi/\partial x$ is replaced by $\delta\phi/\delta x$ where δx is small and y is constant, but the solutions are *not* approximate in the sense of being crude estimates. The data of the problems of technology are invariably subject to errors of measurement, besides which, all arithmetical work is limited to a finite number of significant figures and contains rounding errors, so even analytical solutions provide only approximate numerical answers. Finite-difference methods generally give solutions that are either as accurate as the data warrant or as accurate as is necessary for the technical purposes for which the solutions are required. In both cases a finite-difference solution is as satisfactory as one calculated from an analytical formula. In future, all non-analytical approximation methods will be called numerical methods

They are not of course restricted to problems for which no analytical solutions can be found. The numerical evaluation of an analytical solution is often a laborious task, as can be seen by inspecting the solution of the torsion problem for a rectangular

cross-section defined by $x = \pm a$, $y = \pm b$, namely

$$\phi = b^2 - y^2 - 32b^2\pi^{-3} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \operatorname{sech} \frac{(2n+1)\pi a}{2b} \times \cosh \frac{(2n+1)\pi x}{2b} \cos \frac{(2n+1)\pi y}{2b},$$

and numerical methods generally provide adequate numerical solutions more simply and efficiently. This is certainly so with finite-difference methods for solving partial differential equations.

In these methods (Fig. 1.1), the area of integration of the elliptic equation, i.e. the area S bounded by the closed curve C, is overlayed by a system of rectangular meshes formed by two sets of equally spaced lines, one set parallel to Ox and the other parallel to Oy, and an approximate solution to the differential equation is found at the points of intersection $P_{1,1}, P_{1,2}, \ldots, P_{i,j}, \ldots$ of the parallel lines, which points are called mesh points. (Other terms in common use are pivotal, nodal, grid, or lattice points.) This solution is obtained by approximating the partial differential equation over the area S by n *algebraic* equations involving the values of ϕ at the n mesh points internal to C. The approximation consists of replacing each derivative of the partial differential equation at the point $P_{i,j}$ (say) by a finite-difference approximation in terms of the values of ϕ at $P_{i,j}$ and at neighbouring mesh points and boundary points, and in writing down for each of the n internal mesh points the algebraic equation approximating the differential equation. This process clearly gives n algebraic equations for the n unknowns $\phi_{1,1}$, $\phi_{1,2}, \ldots \phi_{i,j}, \ldots$. Accuracy can usually be improved either by increasing the number of mesh points or by including 'correction terms' in the approximations for the derivatives.

Parabolic and hyperbolic equations

Problems involving time t as one independent variable lead usually to parabolic or hyperbolic equations.

The simplest parabolic equation, $\partial U/\partial t = \kappa \partial^2 U/\partial x^2$, derives from the theory of heat conduction and its solution gives, for example, the temperature U at a distance x units of length from one end of a thermally insulated bar after t seconds of heat

conduction. In such a problem the temperatures at the *ends* of a bar of length l (say) are often known for all time. In other words, the *boundary conditions* are known. It is also usual for the temperature distribution along the bar to be known at some particular instant. This instant is usually taken as zero time and the temperature distribution is called the *initial condition.* The solution gives U for values of x between 0 and l and values of t from zero to infinity. Hence the area of integration S in the x–t plane (Fig. 1.2), is the infinite area bounded by the x-axis and the parallel lines $x = 0, x = l$. This is described as an *open* area because the boundary curves marked C do not constitute a closed boundary in any finite region of the x–t plane.

Applications of finite-difference methods of solution to parabolic equations are no different from their application to elliptic equations in so far as the integration of the differential equation over S is approximated by the solution of algebraic equations. The structure of the algebraic equations is different however in that it propagates the solution forward from one time row to the next in a step-by-step fashion.

Hyperbolic equations generally originate from vibration problems, or from problems where discontinuities can persist in time, such as with shock waves, across which there are discontinuities in speed, pressure and density. The simplest hyperbolic equation is the one-dimensional wave equation $\partial^2 U/\partial t^2 = c^2 \partial^2 U/\partial x^2$, giving, for example, the transverse displacement U at a distance x

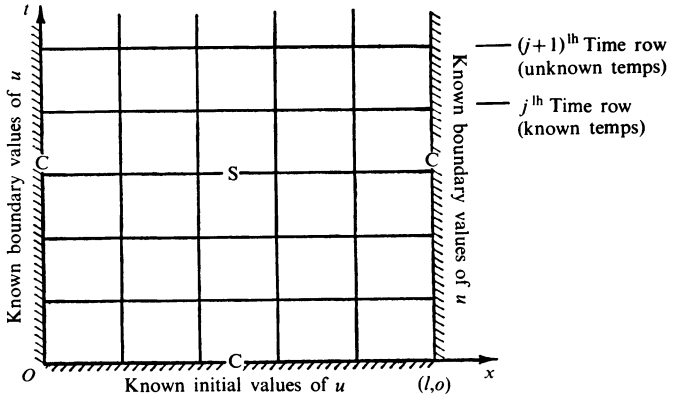

Fig. 1.2

from one end of a vibrating string of length l after a time t. As the values of U at the ends of the string are usually known for all time (the boundary conditions) and the shape and velocity of the string are prescribed at zero time (the initial conditions), it is seen (Fig. 1.2), that the solution is similar to that of a parabolic equation in that the calculation of U for a given x and t, $(0 \leqslant x \leqslant l)$, entails integration of the equation over the open area S bounded by the open curve C. Although hyperbolic equations can be solved numerically by finite-difference methods, those involving only two independent variables, x and t say, are often dealt with by the method of characteristics, especially if the initial conditions and/or boundary conditions involve discontinuities. This method finds special curves in the x–t plane, called characteristic curves, along which the solution of the *partial* differential equation is reduced to the integration of an *ordinary* differential equation. This ordinary equation is generally integrated by numerical methods.

In conclusion, it is worth noting that whereas changes to the shape of the area of integration or to the boundary and initial conditions of partial differential equations often make their analytical solutions impossible, such changes do not fundamentally affect finite-difference methods although they sometimes necessitate rather complicated modifications to the methods.

Finite-difference approximations to derivatives

When a function U and its derivatives are single-valued, finite and continuous functions of x, then by Taylor's theorem,

$$U(x+h) = U(x) + hU'(x) + \tfrac{1}{2}h^2U''(x) + \tfrac{1}{6}h^3U'''(x) + \ldots \tag{1.1}$$

and

$$U(x-h) = U(x) - hU'(x) + \tfrac{1}{2}h^2u''(x) - \tfrac{1}{6}h^3U'''(x) \ldots . \tag{1.2}$$

Addition of these expansions gives

$$U(x+h) + U(x-h) = 2U(x) + h^2U''(x) + O(h^4), \tag{1.3}$$

where $O(h^4)$ denotes terms containing fourth and higher powers of h. Assuming these are negligible in comparison with lower

powers of h it follows that,

$$U''(x) = \left(\frac{d^2U}{dx^2}\right)_{x=x} \simeq \frac{1}{h^2}\{U(x+h) - 2U(x) + U(x-h)\}, \quad (1.4)$$

with a leading error on the right-hand side of order h^2.

Subtraction of eqn (1.2) from eqn (1.1) and neglect of terms of order h^3 leads to

$$U'(x) = \left(\frac{dU}{dx}\right)_{x=x} \simeq \frac{1}{2h}\{U(x+h) - U(x-h)\}, \quad (1.5)$$

with an error of order h^2.

Equation (1.5) clearly approximates the slope of the tangent at P by the slope of the chord AB, and is called a *central-difference* approximation. We can also approximate the slope of the tangent at P by either the slope of the chord PB, giving the *forward-difference* formula,

$$U'(x) \simeq \frac{1}{h}\{U(x+h) - U(x)\}, \quad (1.6)$$

or the slope of the chord AP giving the *backward-difference* formula

$$U'(x) \simeq \frac{1}{h}\{U(x) - U(x-h)\}. \quad (1.7)$$

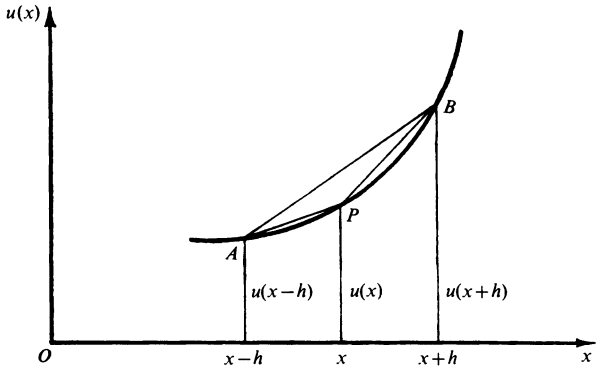

Fig. 1.3

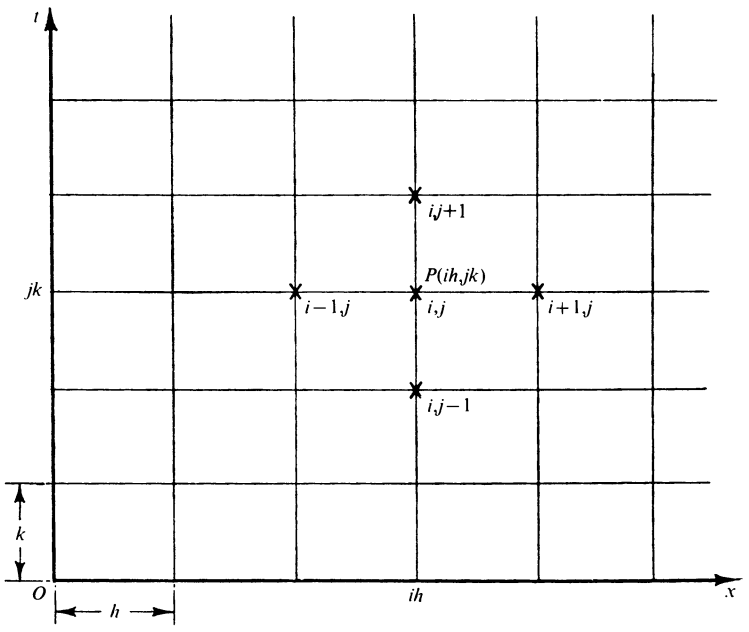

Fig. 1.4

Both (1.6) and (1.7) can be written down immediately from eqns (1.1) and (1.2) respectively, assuming second and higher powers of h are negligible. This shows that the leading errors in these forward and backward-difference formulae are both $O(h)$.

Notation for functions of several variables

Assume U is a function of the independent variables x and t. Subdivide the x–t plane into sets of equal rectangles of sides $\delta x = h$, $\delta t = k$, by equally spaced grid lines parallel to Oy, defined by $x_i = ih$, $i = 0, \pm 1, \pm 2, \ldots$, and equally spaced grid lines parallel to Ox, defined by $t_j = jk$, $j = 0, 1, 2, \ldots$, as shown in Fig. 1.4.

Denote the value of U at the representative mesh point $P(ih, jk)$ by

$$U_P = U(ih, jk) = U_{i,j}.$$

Then by eqn (1.4),

$$\left(\frac{\partial^2 U}{\partial x^2}\right)_P = \left(\frac{\partial^2 U}{\partial x^2}\right)_{i,j} \simeq \frac{U\{(i+1)h, jk\} - 2U\{ih, jk\} + U\{(i-1)h, jk\}}{h^2}.$$

i.e.

$$\frac{\partial^2 U}{\partial x^2} \simeq \frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{h^2}, \tag{1.8}$$

with a leading error of order h^2. Similarly,

$$\left(\frac{\partial^2 U}{\partial t^2}\right)_{i,j} \simeq \frac{U_{i,j+1} - 2U_{i,j} + U_{i,j-1}}{k^2}, \tag{1.9}$$

with a leading error of order k^2.

With this notation the forward-difference approximation for $\partial U/\partial t$ at P is

$$\frac{\partial U}{\partial t} \simeq \frac{U_{i,j+1} - U_{i,j}}{k}, \tag{1.10}$$

with a leading error of $O(k)$.