Cambridge University Press & Assessment
978-1-009-27548-4 — Finite Element and Finite Volume Methods for Heat Transfer and Fluid Dynamics
J. N. Reddy, N. K. Anand, P. Roy
Excerpt
More Information

Part I Preliminaries

1 Mathematical Preliminaries

1.1 Introduction

Most engineering systems can be described, with the aid of the laws of physics and observations, in terms of algebraic, differential, and integral equations. In most problems of practical interest, these equations cannot be solved exactly, mostly because of irregular domains on which the equations are posed, variable coefficients in the equations, complicated boundary conditions, and the presence of nonlinearities. Approximate representation of differential and integral equations to obtain algebraic relations among quantities that characterize the system and implementation of the steps to obtain algebraic equations and their solution using computers constitute a *numerical method*.

The process of converting differential or integral equations to a set of algebraic equations is called *discretization* of the equations. The discretization procedure varies from one numerical method to another. The discretization equations are generated at a finite number of points within the domain. The location and logical numbering of these locations at which dependent variables are calculated is referred to as *grid generation*. The locations at which the dependent variables (unknowns) are calculated are referred to as *grid* points or *nodes*. The terms grid generation and mesh generation are used interchangeably. A valid numerical solution must be independent of the grid (or mesh) size used, and such a solution is called a *grid-independent solution* . If a numerical solution procedure is represented by an equilateral triangle (see Fig. 1.1.1) its vertices would represent: (a) discretization of differential equations, (b) grid generation, and (c) the solution to a system of linear algebraic equations.

There are a number of numerical methods to solve differential equations, the most commonly used methods being the finite difference, finite element, boundary element, and finite volume methods. Because of the power of numerical methods and electronic computation, it is possible to include most relevant features of a physical process in the corresponding mathematical model, without worrying about its exact solution by an analytical means. The finite element method (FEM) [1–6] and the finite volume method (FVM) [7–11] are the most popular and powerful numerical methods devised to analyze many heat transfer and fluid flow problems. While there are numerous books dedicated to each of these two methods, there is only one book, by Chung [12], which deals with both methods as well as the finite difference method (FDM). The main focus of the book by Chung [12] is to introduce the readers all three methods.

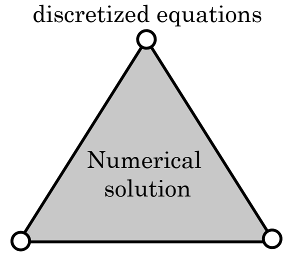

Fig. 1.1.1 Three pillars of numerical solution procedure.

Over the past five decades, computers have made it possible, with the help of suitable mathematical models and numerical methods, to analyze many practical problems of engineering for design and manufacturing. Most of the practicing engineers may end up using a commercial software package or an open source code. But having a background in the numerical methods and having written a computer program based on a numerical method makes them ask the right kind of questions, and be effective users of a commercial software package or an open source code. This book is dedicated to the study of the FEM and FVM as applied to problems of heat transfer and fluid dynamics. The objective is to familiarize the readers with the inner details of these two methods and help them formulate, write their own computer programs, and solve problems of interest to them.

To prepare the readers for an informed walk through various topics that come up in the discussion of differential equations and their solution, we first present some mathematical concepts and definitions. Readers familiar with these topics may skip these and go to Chapter 2.

1.2 Mathematical Models

1.2.1 Preliminary Comments

A set of mathematical relations between variables of a physical system is termed a *mathematical model*. The relationships can be algebraic, differential, and/or integral in nature. Of course, most mathematical models are combinations of algebraic and differential equations. Integral relations usually come through constitutive models (i.e., viscoelastic solids or fluids) and in solving heat transfer problems with radiation. The mathematical models of physical phenomena are often based on fundamental scientific laws of physics such as the principle of conservation of mass, the principle of balance of linear momentum, and the principle of balance of energy (see [13, 14]). The equations resulting from these principles are supplemented by equations that describe the constitutive behavior and boundary and/or initial conditions. A review of the equations of heat transfer and fluid dynamics will be presented in Chapter 2.

1.2.2 Types of Differential Equations

All differential equations can be grouped, based on the number of independent coordinates, into two types: (1) ordinary differential equations (ODEs) are those which contain only one independent coordinate, say x (spatial coordinate) or t (time); and (2) partial differential equations (PDEs) are those which contain two (e.g., x and t or x and y) or more (e.g., x, y, and t; x, y, and z; and x, y, z, and t). Clearly, differential equations describing two- and three-dimensional heat transfer and fluid flow are necessarily PDEs.

Examples of an ODE are [$u = u(t)$ or $u = u(x)$]:

$$a\frac{du}{dt} + bu = f(t), \quad a\frac{d^2u}{dt^2} + c\frac{du}{dt} + pu = f(t) \tag{1.2.1}$$

$$a\frac{du}{dx} + bu = g(x), \quad a\frac{d^2u}{dx^2} + c\frac{du}{dx} + pu = g(x). \tag{1.2.2}$$

Examples of a PDE are:

$$c\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = f(x,t), \quad -\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right) + c\frac{\partial u}{\partial t} + pu = f(x,t) \tag{1.2.3}$$

$$-\frac{\partial}{\partial x}\left(a_{11}\frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(a_{22}\frac{\partial u}{\partial y}\right) = g(x), \quad -\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right) + c\frac{\partial^2 u}{\partial t^2} = g(x,t). \tag{1.2.4}$$

In Eqs. (1.2.1)–(1.2.4), a_{11}, a_{22}, a, b, c, p, f, and g are known functions of x, y, and/or t. The variable u is termed a *dependent variable*.

The *order* of a differential equation is the highest-order derivative appearing in the equation. Thus, an ODE or PDE can be first-, second-, or higher-order. The first equation in (1.2.1), the first equation in (1.2.2), and the first equation (1.2.3) are first-order equations and the others are second-order equations. A PDE can have multiple orders, each with respect to independent coordinates, x, y, z, and t.

An nth-order ODE in a dependent variable $u(\xi)$ contains derivatives of order n and less than n with respect to a coordinate ξ. The solution of an nth-order ODE requires n integrations with respect to the independent coordinate ξ. With each integration, one obtains a constant of integration. To determine the constants of integration, one needs n auxiliary conditions on u. When all conditions on u are specified (or known) at one specific value of the independent coordinate, say $\xi = 0$, we say that the ODE describes an *initial value problem* (IVP) and the conditions are called *initial conditions*. When conditions on u are specified at more than one value of the independent coordinate ξ, then the ODE is said to describe a *boundary value problem* (BVP), and the auxiliary conditions are termed as *boundary conditions*. Thus, a BVP has to be second or higher order or contain a set of first order coupled equations, and all first-order ODEs with a single unknown represent IVPs. There are numerical solution techniques that convert a BVP to an IVP (e.g., shooting methods), but the classification of the original equation as a BVP or an IVP follows the above definition.

The distinction between IVPs and BVPs is an important one as it dictates the way the information (direction and speed) propagates within the domain. The understanding of the propagation of information (solution or disturbance) within the domain dictates the kind of numerical scheme we adopt to get a physically realistic solution. In domains described by IVPs, the information propagates at finite speed in one direction; the independent variable could be space or time. Information in such domains could propagate from the present to the future, or left to right, as illustrated in Fig. 1.2.1(a). Thus, in discretizing a differential operator, one should make sure that the discrertization expression is such that no information comes from the future to the present, if it is a time-dependent problem; or no information comes from downstream to upstream, if it is a space-dependent problem. The IVPs are said to exhibit one-way behavior wherein the information propagates in only one direction. These IVPs are also called *marching problems*. On the other hand, in the BVPs, the information propagates from the boundary points into the domain, as shown in Fig. 1.2.1(b).

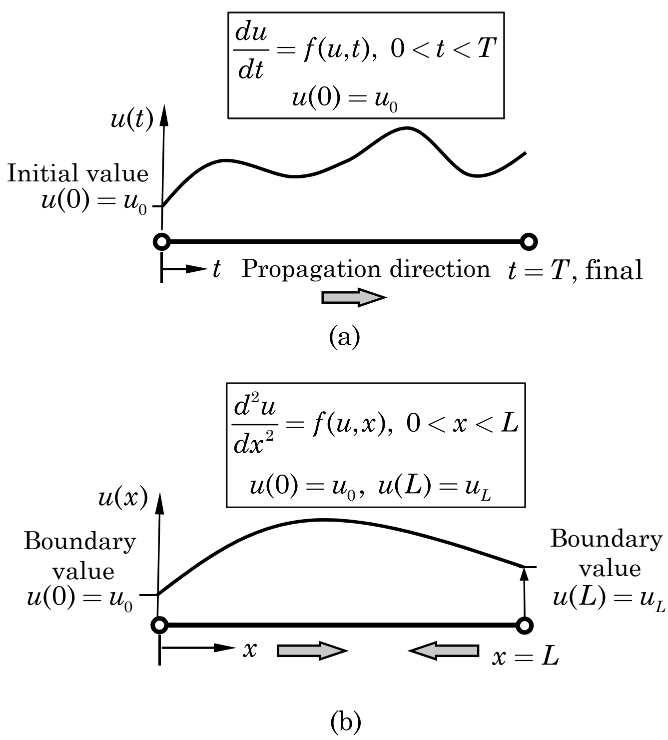

Fig. 1.2.1 (a) Initial value problems (IVPs) and (b) boundary value problems (BVPs) on one-dimensional domains. In IVPs, the initial condition dictates the propagation of the solution into the domain, whereas in the BVPs, the boundary conditions influence the interior solution.

A PDE in the dependent variable $u(\xi, \eta)$, with n derivatives in one independent coordinate ξ and m derivatives in the other independent coordinate η, requires n integrations with respect to ξ and m integrations with respect to η. If the constants that appear due to integration with respect to ξ are determined using known conditions on u and its derivatives at different values of ξ, then the equation describes a BVP with respect to ξ and the conditions are classified as the boundary conditions of the problem. If the constants appearing due to integration with respect to η are determined using known conditions on u and its derivatives at one fixed value (say, at $\eta = 0$), the equation describes an IVP with respect to η and the conditions are classified as initial conditions. When the constants appearing due to integrations with respect to η are also determined using known conditions on u and its derivatives at *different* values of η, the PDE describes a BVP with respect to ξ and η. Thus a PDE can describe a BVP (only) or a boundary value-initial value problem. All time-dependent problems are either pure IVPs or initial value–boundary value problems (IVBPs), when the problems are described by PDEs involving both time and space. However, not all IVPs are time-dependent problems.

In summary, a second-order differential equation in spatial coordinates (ODE or PDE) involves the specification of the dependent variable and its derivatives up to and including order $2n-1$. On the other hand, an mth-order differential equation in time requires the specification of the dependent variable and its derivatives with respect to time of order $m-1$ at the initial time (i.e., $t = 0$). A PDE in space and time can be a BVP as well as an IVP (i.e., it requires both boundary conditions and initial conditions). A first-order ODE in a coordinate ξ requires only one condition, which can be at $\xi = 0$ or $\xi = L$, where L is the final value of ξ.

It is important to keep in mind that the differential equation for a system is only valid within a domain (e.g., $0 < x < L$) and not valid at the boundary points (i.e., at $x = 0$ and $x = L$). Of course, the boundary conditions are valid only at the boundary. However, the solution to boundary value or initial value problems are valid both inside the domain and on the boundary.

An ODE or a PDE is called *nonlinear* when the dependent variable(s) or its derivatives appear in a nonlinear form. A system is said to be linear if it satisfies the *Principle of Superposition* and the *Principle of Proportionality*. All of the ODEs and PDEs presented in Eqs. (1.2.1)–(1.2.4) are linear *unless* the coefficients a_{11}, a_{22}, a, b, c, and p are functions of the dependent variable u and/or its derivatives, or the dependent variable manifests as a nonlinear term. One way to determine the linearity of a differential equation is to write it in an operator form

$$Au = f, \tag{1.2.5}$$

where A is an operator acting on the dependent variable u. For example, the eight operators associated with the eight differential equations in Eqs. (1.2.1)–(1.2.4) are

$$\begin{aligned}
&A_1 u = f, \quad A_1 = a\frac{d}{dt} + b; \quad A_2 u = f, \quad A_2 = a\frac{d^2}{dt^2} + c\frac{d}{dt} + p\\
&A_3 u = g, \quad A_3 = a\frac{d}{dx} + b; \quad A_4 u = g, \quad A_4 = a\frac{d^2}{dx^2} + c\frac{d}{dx} + p\\
&\qquad\qquad A_5 u = f, \quad A_5 = c\frac{\partial}{\partial t} + \frac{\partial}{\partial x}
\end{aligned} \tag{1.2.6a}$$

$$
\begin{aligned}
A_6 u = f, &\quad A_6 = -\frac{\partial}{\partial x}\left(a\frac{\partial}{\partial x}\right) + c\frac{\partial}{\partial t} + p \\
A_7 u = g, &\quad A_7 = -\frac{\partial}{\partial x}\left(a_{11}\frac{\partial}{\partial x}\right) - \frac{\partial}{\partial y}\left(a_{22}\frac{\partial}{\partial y}\right) \\
A_8 u = g, &\quad A_8 = -\frac{\partial}{\partial x}\left(a\frac{\partial}{\partial x}\right) + c\frac{\partial^2}{\partial t^2}.
\end{aligned}
\tag{1.2.6b}
$$

An operator A is said to be linear, hence the differential equation (1.2.5) is said to be a linear differential equation, if the following equality [13],

$$A(\alpha u + \beta v) = \alpha A(u) + \beta A(v) \tag{1.2.7}$$

holds for all real numbers α and β. Otherwise, the operator is nonlinear, and the associated differential equation is a nonlinear differential equation. Examples of nonlinear operators are provided by

$$A_1 = \frac{d^2}{dx^2} + cu\frac{d}{dx} + p, \quad A_2 = -\frac{\partial}{\partial x}\left(\sqrt{u}\frac{\partial}{\partial x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\frac{\partial}{\partial y}\right). \tag{1.2.8}$$

All differential equations have two kinds of terms: (1) terms involving the dependent variables and their derivatives of certain order, and (2) terms that contain only the independent variable(s). Of course, the terms containing the dependent variables can also have coefficients that are functions of the independent variables (as well as the dependent variables, when the differential equations are nonlinear). The terms involving only the independent variables (i.e., x, y, and/or t) are typically known as the *source terms*. When the source term in a differential equation is zero, the equation is called a *homogeneous differential equation*. Referring to Eqs. (1.2.1)–(1.2.4), the functions f and g (of x, y, and/or t) are the source terms. Similarly, when a specified initial value or boundary value is nonzero, it is said to be a nonhomogeneous initial or boundary condition. Otherwise, the initial or boundary conditions is called a homogeneous initial or boundary condition. The other way to look at this concept is that, if a dependent variable or its derivative appears in every term of a differential equation or of an initial condition or of a boundary condition then they are referred to as *homogeneous* equations or auxiliary (initial or boundary) conditions respectively. The phrase *general solution* to a differential equation means it is the solution of a homogeneous differential equation. The part of the total solution to a differential equation that satisfies the nonhomogeneous differential equation is known as the *particular solution*.

1.2.3 Examples of Mathematical Models

1.2.3.1 A Boundary Value Problem: the Heat Conduction Problem

As an example of a mathematical model, we consider one-dimensional heat flow in an uninsulated circular cylindrical rod (see Fig. 1.2.2):

$$-\frac{d}{dx}\left(Ak\frac{dT}{dx}\right) + \beta P(T - T_\infty) = g, \tag{1.2.9}$$

where T denotes temperature (°C) above a certain reference temperature, A is the area of cross section (m^2) of the rod, P is the perimeter (m), k is the conductivity [W/(m ·°C)], β is the convective heat transfer coefficient [W/(m^2·°C)], and g is the internal heat generation per unit length (W/m). Equation (1.2.9) is a statement of balance of energy [15]. The first term denotes the transfer of energy due to conduction (diffusion) and the second term is the energy transfer due to convection through the surface of the rod.

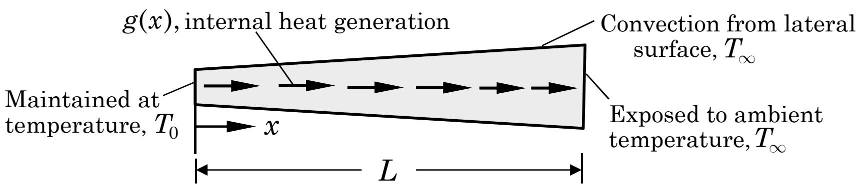

Fig. 1.2.2 One-dimensional heat flow in a rod with internal heat generation $g(x)$.

Equation (1.2.9) is a linear, nonhomogeneous (because the right-hand side is nonzero), second-order differential equation with variable coefficients (because, in general, kA can be a function of x), which can be solved with two known conditions on either temperature T or heat $Q = -kA(dT/dx)$ (but not both) at each of the two boundary points (two ends of the bar).

Suppose that the auxiliary conditions at the boundary points $x = 0$ and $x = L$ are of the form

$$T(0) = T_0, \quad \left[kA\frac{dT}{dx} + \beta A(T - T_\infty)\right]_{x=L} = Q_0, \tag{1.2.10}$$

where T_0 and Q_0 are the specified temperature and specified heat, respectively. These auxiliary conditions are, clearly, boundary conditions. The first condition in Eq. (1.2.10) is called the *Dirichlet* boundary condition. The second condition represents the balance of heat due to conduction $[kA(dT/dx)]$, and convection $[\beta A(T - T_\infty)]$ at $x = L$, and it is known as the *Newton*, *mixed*, or *Robin* boundary condition. In the heat transfer literature, it is also known as the *convective* boundary condition. As a special case, the boundary condition when the end $x = L$ is insulated is given by setting $\beta = 0$ and $Q_0 = 0$ in Eq. (1.2.10). The case where $\beta = 0$ is known as the *Neumann* boundary condition.

Equation (1.2.9) is an ODE because it contains only one independent coordinate, namely, x. Also, Eq. (1.2.9) is a *nonhomogeneous differential equation* because the source term (g) is nonzero. We also note that the boundary conditions in Eq. (1.2.10) are nonhomogeneous, as long as at least one of the three quantities (T_0, T_∞, Q_0) is nonzero.

The problem described by Eqs. (1.2.9) and (1.2.10) is a BVP because its solution requires the specification of the auxiliary conditions at more than one specific value of the independent variable (i.e., $x = 0$ and $x = L$). Quantities k, A, β, P, T_∞, T_0, Q_0, and f are called the *data* or *parameters* of the problem because they are prescribed quantities of the problem.

1.2.3.2 An Initial Value Problem: the Simple Pendulum Problem

Next, consider a simple pendulum, which consists of a bob of constant mass m (kg) attached to one end of a rod of length ℓ (m) with the other end pivoted to a fixed point O (without friction), as shown in Fig. 1.2.3. The bob and the rod are assumed to be rigid (i.e., not deformable) and the rod is massless.

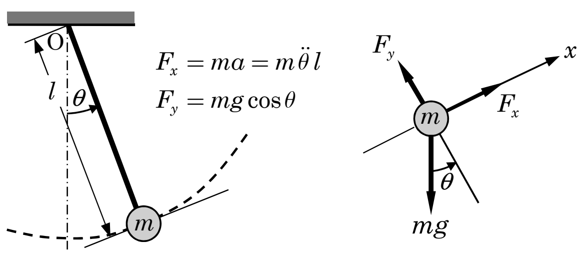

Fig. 1.2.3 Simple pendulum (a bob of mass m is attached to a massless rod of length l), which is connected to point O and free to rotate about the point O.

The equation governing the motion of the simple pendulum can be determined using Newton's Second Law of motion (i.e., the vector sum of externally applied forces on a system is equal to the time rate of change of the linear momentum of the system):

$$m\frac{d^2\theta}{dt^2} + \frac{mg}{\ell}\sin\theta = 0, \tag{1.2.11}$$

where θ is the angular displacement (radians), g is the acceleration due to gravity (m/s^2), and t denotes time (s).

Equation (1.2.11) is nonlinear (in θ) on account of the term $\sin\theta$. As explained before, a check for linearity is to replace the dependent unknown (θ) with a constant multiple of itself and see if the constant can be factored out. For example, replacing θ with $\alpha\theta$ in Eq. (1.2.11), we obtain $m\alpha\frac{d^2\theta}{dt^2} + \frac{mg}{\ell}\sin(\alpha\theta) = 0$. Although α is factored out in the first term (hence the first term is linear), it cannot be factored out in the second term (hence the second term is nonlinear); thus, the whole equation is said to be nonlinear. For small angular motions, $\sin\theta$ can be approximated as $\sin\theta \approx \theta$. Then, Eq. (1.2.3) becomes a homogeneous, second-order, linear ODE (the independent coordinate being t):

$$m\frac{d^2\theta}{dt^2} + \frac{mg}{\ell}\theta = 0. \tag{1.2.12}$$

The solution of Eq. (1.2.11) or Eq. (1.2.12) requires knowledge of conditions on θ and its time derivative $\dot{\theta}$ (angular velocity) at time $t = 0$. These conditions are the initial conditions. Thus, the linear problem involves solving the second-order differential equation (1.2.12) subjected to the initial conditions:

$$\theta(0) = \theta_0, \qquad \left.\frac{d\theta}{dt}\right|_{t=0} = v_0. \tag{1.2.13}$$

Since the problem described by Eqs. (1.2.12) and (1.2.13) requires two known conditions at the same value of the independent coordinate (t), it is an IVP. The initial conditions are said to be homogeneous when both θ_0 and v_0 are zero.

If we wish to solve the nonlinear equation, Eq. (1.2.11), subject to the conditions in Eq. (1.2.13), we may consider using a numerical method because it is not possible to solve Eq. (1.2.11) exactly for large values of θ.

1.2.4 Numerical Solution of First-Order Ordinary Differential Equations

1.2.4.1 The Euler Methods

Let us consider the following first-order ODE in $y(x)$ with an end condition:

$$\frac{dy}{dx} = f(x, y), \quad y(0) = y_0. \tag{1.2.14}$$

Integrating Eq. (1.2.14) over the interval $[x_i, x_{i+1}]$, we obtain

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x, y)dx, \tag{1.2.15}$$

where $y_{i+1} = y(x_{i+1})$ and so on.

One can complete the integration on the right-hand side if one assumes f to remain constant in the interval $[x_i, x_{i+1}]$. But the question is where to evaluate f in this interval when it is a function of x and y (see Fig. 1.2.4). The stability of the method and the complexity of the calculation depend on where $f(x, y)$ or the slope is evaluated in this interval. If the $f(x, y)$ is evaluated at the initial point, then

$$y_{i+1} = y_i + hf(x_i, y_i), \quad h = x_{i+1} - x_i. \tag{1.2.16}$$

In the above equation, an unknown (y_{i+1}) is evaluated in terms of all known quantities. Such a formulation is referred to as an *explicit method*. If $f(x, y)$ is evaluated at the end point x_{i+1}, then we obtain

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}); \quad \text{for } i = 0 \text{ we have } \ y_1 = y_0 + hf(x_0, y_0). \tag{1.2.17}$$

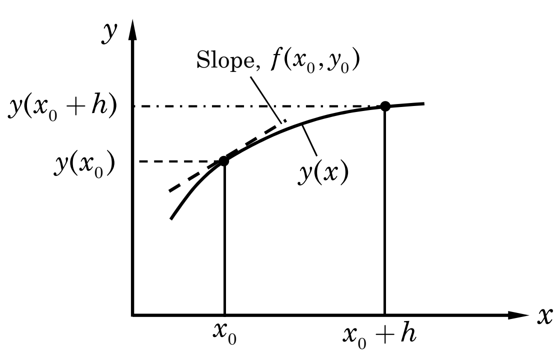

Fig. 1.2.4 Values of a function $y(x)$ at different points, $x = x_0$ and $x = x_0 + h$.