8.2

FINITE DIFFERENCE, FINITE ELEMENT AND FINITE VOLUME METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS

Joaquim Peiró and Spencer Sherwin
Department of Aeronautics, Imperial College, London, UK

There are three important steps in the computational modelling of any physical process: (i) problem definition, (ii) mathematical model, and (iii) computer simulation.

The first natural step is to define an idealization of our problem of interest in terms of a set of relevant quantities which we would like to measure. In defining this idealization we expect to obtain a well-posed problem, this is one that has a unique solution for a given set of parameters. It might not always be possible to guarantee the fidelity of the idealization since, in some instances, the physical process is not totally understood. An example is the complex environment within a nuclear reactor where obtaining measurements is difficult.

The second step of the modeling process is to represent our idealization of the physical reality by a mathematical model: the governing equations of the problem. These are available for many physical phenomena. For example, in fluid dynamics the Navier–Stokes equations are considered to be an accurate representation of the fluid motion. Analogously, the equations of elasticity in structural mechanics govern the deformation of a solid object due to applied external forces. These are complex general equations that are very difficult to solve both analytically and computationally. Therefore, we need to introduce simplifying assumptions to reduce the complexity of the mathematical model and make it amenable to either exact or numerical solution. For example, the irrotational (without vorticity) flow of an incompressible fluid is accurately represented by the Navier–Stokes equations but, if the effects of fluid viscosity are small, then Laplace's equation of *potential flow* is a far more efficient description of the problem.

S. Yip (ed.),
Handbook of Materials Modeling. Volume I: Methods and Models, 1–32.

After the selection of an appropriate mathematical model, together with suitable boundary and initial conditions, we can proceed to its solution. In this chapter we will consider the numerical solution of mathematical problems which are described by partial differential equations (PDEs). The three classical choices for the numerical solution of PDEs are the finite difference method (FDM), the finite element method (FEM) and the finite volume method (FVM).

The FDM is the oldest and is based upon the application of a local Taylor expansion to approximate the differential equations. The FDM uses a topologically square network of lines to construct the discretization of the PDE. This is a potential bottleneck of the method when handling complex geometries in multiple dimensions. This issue motivated the use of an integral form of the PDEs and subsequently the development of the finite element and finite volume techniques.

To provide a short introduction to these techniques we shall consider each type of discretization as applied to one-dimensional PDEs. This will not allow us to illustrate the geometric flexibility of the FEM and the FVM to their full extent, but we will be able to demonstrate some of the similarities between the methods and thereby highlight some of the relative advantages and disadvantages of each approach. For a more detailed understanding of the approaches we refer the reader to the section on suggested reading at the end of the chapter.

The section is structured as follows. We start by introducing the concept of conservation laws and their differential representation as PDEs and the alternative integral forms. We next discusses the classification of partial differential equations: elliptic, parabolic, and hyperbolic. This classification is important since the type of PDE dictates the form of boundary and initial conditions required for the problem to be well-posed. It also, permits in some cases, e.g., in hyperbolic equations, to identify suitable schemes to discretise the differential operators. The three types of discretisation: FDM, FEM and FVM are then discussed and applied to different types of PDEs. We then end our overview by discussing the numerical difficulties which can arise in the numerical solution of the different types of PDEs using the FDM and provides an introduction to the assessment of the stability of numerical schemes using a Fourier or Von Neumann analysis.

Finally we note that, given the scientific background of the authors, the presentation has a bias towards fluid dynamics. However, we stress that the fundamental concepts presented in this chapter are generally applicable to continuum mechanics, both solids and fluids.

1. Conservation Laws: Integral and Differential Forms

The governing equations of continuum mechanics representing the kinematic and mechanical behaviour of general bodies are commonly referred

to as *conservation laws*. These are derived by invoking the conservation of mass and energy and the momentum equation (Newton's law). Whilst they are equally applicable to solids and fluids, their differing behaviour is accounted for through the use of a different constitutive equation.

The general principle behind the derivation of conservation laws is that the rate of change of $u(\boldsymbol{x}, t)$ within a volume V plus the flux of u through the boundary A is equal to the rate of production of u denoted by $S(u, \boldsymbol{x}, t)$. This can be written as

$$\frac{\partial}{\partial t}\int_V u(\boldsymbol{x}, t)\,\mathrm{d}V + \int_A \boldsymbol{f}(u)\cdot\boldsymbol{n}\,\mathrm{d}A - \int_V S(u, \boldsymbol{x}, t)\,\mathrm{d}V = 0 \tag{1}$$

which is referred to as the *integral* form of the conservation law. For a fixed (independent of t) volume and, under suitable conditions of smoothness of the intervening quantities, we can apply Gauss' theorem

$$\int_V \nabla\cdot\boldsymbol{f}\,\mathrm{d}V = \int_A \boldsymbol{f}\cdot\boldsymbol{n}\,\mathrm{d}A$$

to obtain

$$\int_V \left(\frac{\partial u}{\partial t} + \nabla\cdot\boldsymbol{f}(u) - S\right)\mathrm{d}V = 0. \tag{2}$$

For the integral expression to be zero for any volume V, the integrand must be zero. This results in the *strong* or differential form of the equation

$$\frac{\partial u}{\partial t} + \nabla\cdot\boldsymbol{f}(u) - S = 0. \tag{3}$$

An alternative *integral* form can be obtained by the method of weighted residuals. Multiplying Eq. (3) by a *weight* function $w(\boldsymbol{x})$ and integrating over the volume V we obtain

$$\int_V \left(\frac{\partial u}{\partial t} + \nabla\cdot\boldsymbol{f}(u) - S\right) w(\boldsymbol{x})\,\mathrm{d}V = 0. \tag{4}$$

If Eq. (4) is satisfied for any weight function $w(\boldsymbol{x})$, then Eq. (4) is equivalent to the differential form (3). The smoothness requirements on $\boldsymbol{f}$ can be relaxed by applying the Gauss' theorem to Eq. (4) to obtain

$$\int_V \left[\left(\frac{\partial u}{\partial t} - S\right) w(\boldsymbol{x}) - \boldsymbol{f}(u)\cdot\nabla w(\boldsymbol{x})\right]\mathrm{d}V + \int_A \boldsymbol{f}\cdot\boldsymbol{n}\, w(\boldsymbol{x})\,\mathrm{d}A = 0. \tag{5}$$

This is known as the *weak* form of the conservation law.

Although the above formulation is more commonly used in fluid mechanics, similar formulations are also applied in structural mechanics. For instance, the well-known principle of virtual work for the static equilibrium of a body [1], is given by

$$\delta W = \int_V (\nabla \boldsymbol{\sigma} + \boldsymbol{f}) \cdot \delta \boldsymbol{v} \, \mathrm{d}V = 0$$

where δW denotes the virtual work done by an arbitrary virtual velocity $\delta \boldsymbol{v}$, $\boldsymbol{\sigma}$ is the stress tensor and $\boldsymbol{f}$ denotes the body force. The similarity with the method of weighted residuals (4) is evident.

2. Model Equations and their Classification

In the following we will restrict ourselves to the analysis of one-dimensional conservation laws representing the transport of a scalar variable $u(x,t)$ defined in the domain $\Omega = \{x, t : 0 \leq x \leq 1, 0 \leq t \leq T\}$. The convection–diffusion-reaction equation is given by

$$\mathcal{L}(u) = \frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(au - b\frac{\partial u}{\partial x}\right) - r\,u = s \tag{6}$$

together with appropriate boundary conditions at $x = 0$ and 1 to make the problem well-posed. In the above equation $\mathcal{L}(u)$ simply represents a linear differential operator. This equation can be recast in the form (3) with $f(u) = au - \partial u/\partial x$ and $S(u) = s + ru$. It is linear if the coefficient a, b, r and s are functions of x and t, and non-linear if any of them depends on the solution, u.

In what follows, we will use for convenience the convention that the presence of a subscript x or t under a expression indicates a derivative or partial derivative with respect to this variable, for example

$$u_x(x) = \frac{\mathrm{d}u}{\mathrm{d}x}(x); \qquad u_t(x,t) = \frac{\partial u}{\partial t}(x,t); \qquad u_{xx}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t).$$

Using this notation, Eq. (6) is re-written as

$$u_t + (au - bu_x)_x - ru = s.$$

2.1. Elliptic Equations

The steady-state solution of Eq. (6) when advection and source terms are neglected, i.e., a=0 and s=0, is a function of x only and satisfies the Helmholtz equation

$$(bu_x)_x + ru = 0. \tag{7}$$

This equation is elliptic and its solution depends on two families of integration constants that are fixed by prescribing boundary conditions at the ends of the domain. One can either prescribe Dirichlet boundary conditions at both ends, e.g., $u(0) = \alpha_0$ and $u(1) = \alpha_1$, or substitute one of them (or both if $r \neq 0$) by a Neumann boundary condition, e.g., $u_x(0) = g$. Here α_0, α_1 and g are known constant values. We note that if we introduce a perturbation into a Dirichlet boundary condition, e.g., $u(0) = \alpha_0 + \epsilon$, we will observe an instantaneous modification to the solution throughout the domain. This is indicative of the elliptic nature of the problem.

2.2. Parabolic Equations

Taking $a = 0$, $r = 0$ and $s = 0$ in our model, Eq. (6) leads to the heat or diffusion equation

$$u_t - (b\,u_x)_x = 0, \tag{8}$$

which is parabolic. In addition to appropriate boundary conditions of the form used for elliptic equations, we also require an initial condition at $t = 0$ of the form $u(x, 0) = u_0(x)$ where u_0 is a given function.

If b is constant, this equation admits solutions of the form $u(x, t) = Ae^{\beta t} \sin kx$ if $\beta + k^2 b = 0$. A notable feature of the solution is that it decays when b is positive as the exponent $\beta < 0$. The rate of decay is a function of b. The more diffusive the equation (i.e., larger b) the faster the decay of the solution is. In general the solution can be made up of many sine waves of different frequencies, i.e., a Fourier expansion of the form

$$u(x, t) = \sum_m A_m \mathrm{e}^{\beta_m t} \sin\, k_m x,$$

where A_m and k_m represent the amplitude and the frequency of a Fourier mode, respectively. The decay of the solution depends on the Fourier contents of the initial data since $\beta_m = -k_m^2 b$. High frequencies decay at a faster rate than the low frequencies which physically means that the solution is being smoothed. This is illustrated in Fig. 1 which shows the time evolution of $u(x, t)$ for an initial condition $u_0(x) = 20x$ for $0 \leq x \leq 1/2$ and $u_0(x) = 20(1 - x)$ for $1/2 \leq x \leq 1$. The solution shows a rapid smoothing of the slope discontinuity of the initial condition at $x = 1/2$. The presence of a positive diffusion ($b > 0$) physically results in a smoothing of the solution which stabilizes it. On the other hand, negative diffusion ($b < 0$) is de-stabilizing but most physical problems have positive diffusion.

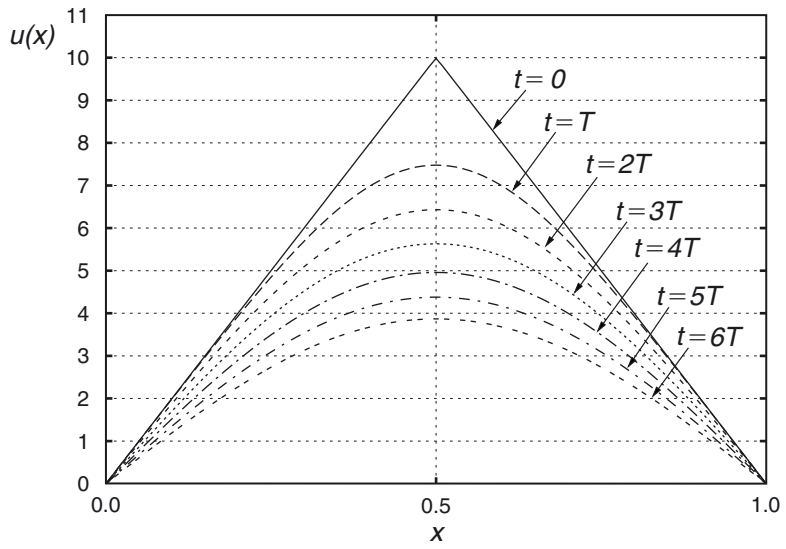

Figure 1. Rate of decay of the solution to the diffusion equation.

2.3. Hyperbolic Equations

A classic example of hyperbolic equation is the linear advection equation

$$u_t + a\,u_x = 0, \tag{9}$$

where a represents a constant velocity. The above equation is also clearly equivalent to Eq. (6) with $b = r = s = 0$. This hyperbolic equation also requires an initial condition, $u(x, 0) = u_0(x)$. The question of what boundary conditions are appropriate for this equation can be more easily be answered after considering its solution. It is easy to verify by substitution in (9) that the solution is given by $u(x, t) = u_0(x - at)$. This describes the propagation of the quantity $u(x, t)$ moving with speed "a" in the x-direction as depicted in Fig. 2. The solution is constant along the *characteristic line* $x - at = c$ with $u(x, t) = u_0(c)$.

From the knowledge of the solution, we can appreciate that for $a > 0$ a boundary condition should be prescribed at $x = 0$, (e.g., $u(0) = \alpha_0$) where information is being fed into the solution domain. The value of the solution at $x = 1$ is determined by the initial conditions or the boundary condition at $x = 0$ and cannot, therefore, be prescribed. This simple argument shows that, in a hyperbolic problem, the selection of appropriate conditions at a boundary point depends on the solution at that point. If the velocity is negative, the previous treatment of the boundary conditions is reversed.

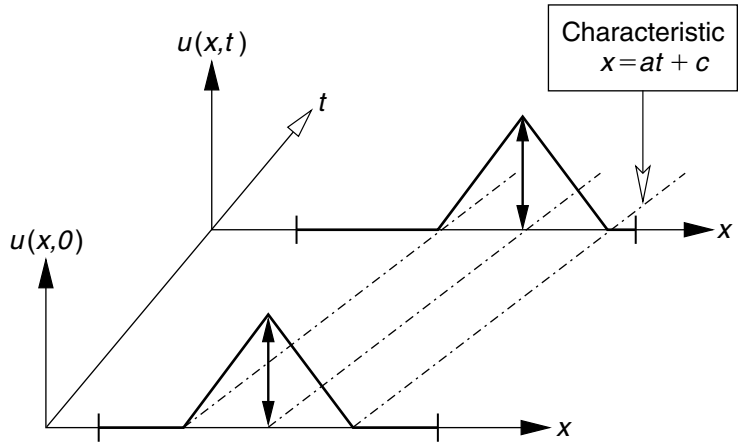

Figure 2. Solution of the linear advection equation.

The propagation velocity can also be a function of space, i.e., $a = a(x)$ or even the same as the quantity being propagated, i.e., $a = u(x, t)$. The choice $a = u(x, t)$ leads to the non-linear inviscid Burgers' equation

$$u_t + u\, u_x = 0. \tag{10}$$

An analogous analysis to that used for the advection equation shows that $u(x, t)$ is constant if we are moving with a local velocity also given by $u(x, t)$. This means that some regions of the solution advance faster than other regions leading to the formation of sharp gradients. This is illustrated in Fig. 3. The initial velocity is represented by a triangular "zig-zag" wave. Peaks and troughs in the solution will advance, in opposite directions, with maximum speed. This will eventually lead to an overlap as depicted by the dotted line in Fig. 3. This results in a non-uniqueness of the solution which is obviously non-physical and to resolve this problem we must allow for the formation and propagation of discontinuities when two characteristics intersect (see Ref. [2] for further details).

3. Numerical Schemes

There are many situations where obtaining an exact solution of a PDE is not possible and we have to resort to approximations in which the infinite set of values in the continuous solution is represented by a finite set of values referred to as the *discrete* solution.

For simplicity we consider first the case of a function of one variable $u(x)$. Given a set of points x_i; $i = 1, \ldots, N$ in the domain of definition of $u(x)$, as

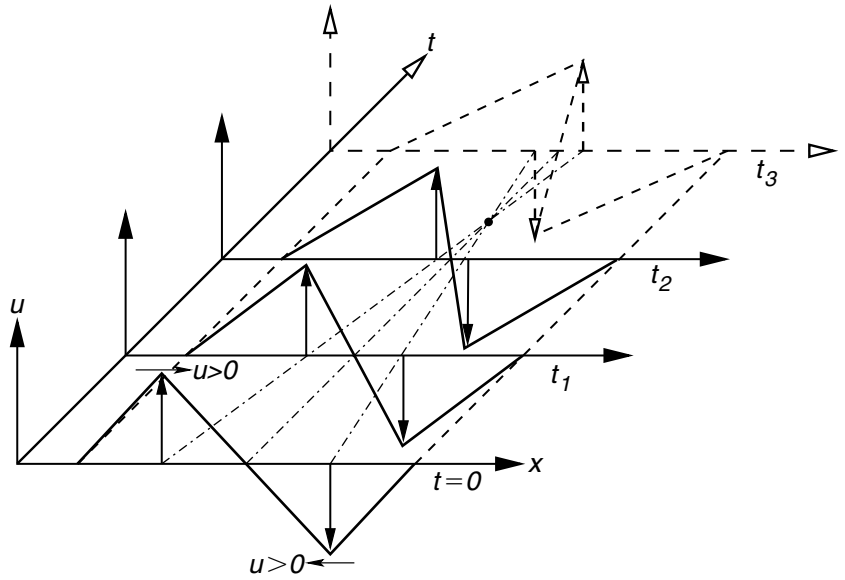

Figure 3. Formation of discontinuities in the Burgers' equation.

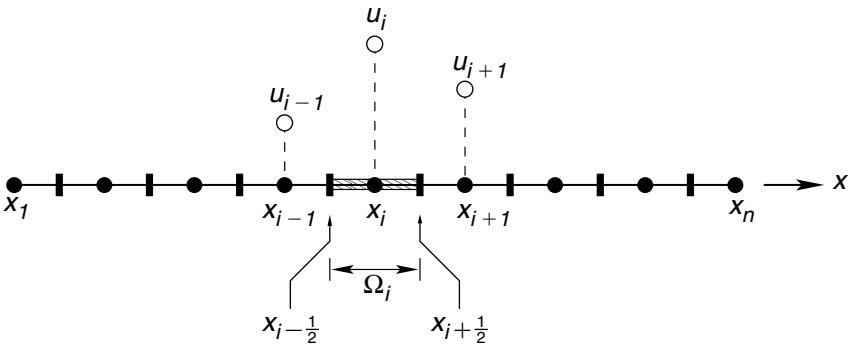

Figure 4. Discretization of the domain.

shown in Fig. 4, the numerical solution that we are seeking is represented by a discrete set of function values $\{u_1, \ldots, u_N\}$ that approximate u at these points, i.e., $u_i \approx u(x_i)$; $i = 1, \ldots, N$.

In what follows, and unless otherwise stated, we will assume that the points are equally spaced along the domain with a constant distance $\Delta x = x_{i+1} - x_i$; $i = 1, \ldots, N-1$. This way we will write $u_{i+1} \approx u(x_{i+1}) = u(x_i + \Delta x)$. This partition of the domain into smaller subdomains is referred to as a *mesh* or *grid*.

3.1. The Finite Difference Method (FDM)

This method is used to obtain numerical approximations of PDEs written in the strong form (3). The derivative of $u(x)$ with respect to x can be defined as

$$\begin{aligned} u_x|_i = u_x(x_i) &= \lim_{\Delta x \to 0} \frac{u(x_i + \Delta x) - u(x_i)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{u(x_i) - u(x_i - \Delta x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{u(x_i + \Delta x) - u(x_i - \Delta x)}{2\Delta x}. \end{aligned} \tag{11}$$

All these expressions are mathematically equivalent, i.e., the approximation converges to the derivative as $\Delta x \to 0$. If Δx is small but finite, the various terms in Eq. (11) can be used to obtain approximations of the derivate u_x of the form

$$u_x|_i \approx \frac{u_{i+1} - u_i}{\Delta x} \tag{12}$$

$$u_x|_i \approx \frac{u_i - u_{i-1}}{\Delta x} \tag{13}$$

$$u_x|_i \approx \frac{u_{i+1} - u_{i-1}}{2\Delta x}. \tag{14}$$

The expressions (12)–(14) are referred to as forward, backward and centred finite difference approximations of $u_x|_i$, respectively. Obviously these approximations of the derivative are different.

3.1.1. Errors in the FDM

The analysis of these approximations is performed by using Taylor expansions around the point x_i. For instance an approximation to u_{i+1} using $n + 1$ terms of a Taylor expansion around x_i is given by

$$\begin{aligned} u_{i+1} = u_i + u_x|_i \, \Delta x + u_{xx}|_i \, \frac{\Delta x^2}{2} + \cdots + \left.\frac{\mathrm{d}^n u}{\mathrm{d}x^n}\right|_i \frac{\Delta x^n}{n!} \\ + \underline{\frac{\mathrm{d}^{n+1} u}{\mathrm{d}x^{n+1}}(x^*) \frac{\Delta x^{n+1}}{(n+1)!}}. \end{aligned} \tag{15}$$

The underlined term is called the remainder with $x_i \leq x^* \leq x_{i+1}$, and represents the error in the approximation if only the first n terms in the expansion are kept. Although the expression (15) is exact, the position x^* is unknown.

To illustrate how this can be used to analyse finite difference approximations, consider the case of the forward difference approximation (12) and use the expansion (15) with $n = 1$ (two terms) to obtain

$$\frac{u_{i+1} - u_i}{\Delta x} = u_x|_i + \frac{\Delta x}{2} u_{xx}(x^*). \tag{16}$$

We can now write the approximation of the derivative as

$$u_x|_i = \frac{u_{i+1} - u_i}{\Delta x} + \epsilon_{\mathrm{T}} \tag{17}$$

where ϵ_{T} is given by

$$\epsilon_{\mathrm{T}} = -\frac{\Delta x}{2} u_{xx}(x^*). \tag{18}$$

The term ϵ_{T} is referred to as the *truncation error* and is defined as the difference between the exact value and its numerical approximation. This term depends on Δx but also on u and its derivatives. For instance, if $u(x)$ is a linear function then the finite difference approximation is exact and $\epsilon_{\mathrm{T}} = 0$ since the second derivative is zero in (18).

The *order* of a finite difference approximation is defined as the power p such that $\lim_{\Delta x \to 0}(\epsilon_{\mathrm{T}}/\Delta x^p) = \gamma \neq 0$, where γ is a finite value. This is often written as $\epsilon_{\mathrm{T}} = O(\Delta x^p)$. For instance, for the forward difference approximation (12), we have $\epsilon_{\mathrm{T}} = O(\Delta x)$ and it is said to be first-order accurate ($p = 1$).

If we apply this method to the backward and centred finite difference approximations (13) and (14), respectively, we find that, for constant Δx, their errors are

$$u_x|_i = \frac{u_i - u_{i-1}}{\Delta x} + \frac{\Delta x}{2} u_{xx}(x^*) \Rightarrow \epsilon_{\mathrm{T}} = O(\Delta x) \tag{19}$$

$$u_x|_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x} - \frac{\Delta x^2}{12} u_{xxx}(x^\star) \Rightarrow \epsilon_{\mathrm{T}} = O(\Delta x^2) \tag{20}$$

with $x_{i-1} \leq x^* \leq x_i$ and $x_{i-1} \leq x^\star \leq x_{i+1}$ for Eqs. (19) and (20), respectively.

This analysis is confirmed by the numerical results presented in Fig. 5 that displays, in logarithmic axes, the exact and truncation errors against Δx for the backward and the centred finite differences. Their respective truncation errors ϵ_{T} are given by (19) and (20) calculated here, for lack of a better value, with $x^* = x^\star = x_i$. The exact error is calculated as the difference between the exact value of the derivative and its finite difference approximation.

The slope of the lines are consistent with the order of the truncation error, i.e., 1:1 for the backward difference and 1:2 for the centred difference. The discrepancies between the exact and the numerical results for the smallest values of Δx are due to the use of finite precision computer arithmetic or round-off error. This issue and its implications are discussed in more detail in numerical analysis textbooks as in Ref. [3].

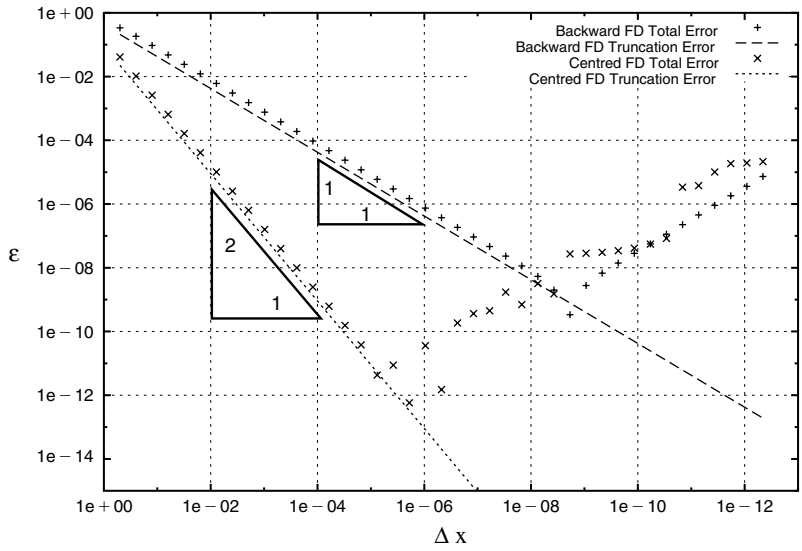

Figure 5. Truncation and rounding errors in the finite difference approximation of derivatives.

3.1.2. Derivation of approximations using Taylor expansions

The procedure described in the previous section can be easily transformed into a general method for deriving finite difference schemes. In general, we can obtain approximations to higher order derivatives by selecting an appropriate number of interpolation points that permits us to eliminate the highest term of the truncation error from the Taylor expansions. We will illustrate this with some examples. A more general description of this derivation can be found in Hirsch (1988).

A second-order accurate finite difference approximation of the derivative at x_i can be derived by considering the values of u at three points: x_{i-1}, x_i and x_{i+1}. The approximation is constructed as a weighted average of these values $\{u_{i-1}, u_i, u_{i+1}\}$ such as

$$u_x|_i \approx \frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x}. \tag{21}$$

Using Taylor expansions around x_i we can write

$$u_{i+1} = u_i + \Delta x\ u_x|_i + \frac{\Delta x^2}{2}\ u_{xx}|_i + \frac{\Delta x^3}{6}\ u_{xxx}|_i + \cdots \tag{22}$$

$$u_{i-1} = u_i - \Delta x\ u_x|_i + \frac{\Delta x^2}{2}\ u_{xx}|_i - \frac{\Delta x^3}{6}\ u_{xxx}|_i + \cdots \tag{23}$$

Putting (22), (23) into (21) we get

$$\frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x} = (\alpha + \beta + \gamma)\frac{1}{\Delta x}u_i + (\alpha - \gamma)\, u_x|_i + (\alpha + \gamma)\frac{\Delta x}{2}\, u_{xx}|_i + (\alpha - \gamma)\frac{\Delta x^2}{6}\, u_{xxx}|_i + (\alpha + \gamma)\frac{\Delta x^3}{12}\, u_{xxxx}|_i + O(\Delta x^4) \tag{24}$$

We require three independent conditions to calculate the three unknowns α, β and γ. To determine these we impose that the expression (24) is consistent with increasing orders of accuracy. If the solution is constant, the left-hand side of (24) should be zero. This requires the coefficient of $(1/\Delta x)u_i$ to be zero and therefore $\alpha+\beta+\gamma = 0$. If the solution is linear, we must have $\alpha-\gamma = 1$ to match $u_x|_i$. Finally whilst the first two conditions are necessary for consistency of the approximation in this case we are free to choose the third condition. We can therefore select the coefficient of $(\Delta x/2)\, u_{xx}|_i$ to be zero to improve the accuracy, which means $\alpha + \gamma = 0$.

Solving these three equations we find the values $\alpha = 1/2$, $\beta = 0$ and $\gamma = -(1/2)$ and recover the second-order accurate centred formula

$$u_x|_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x} + O(\Delta x^2).$$

Other approximations can be obtained by selecting a different set of points, for instance, we could have also chosen three points on the side of x_i, e.g., u_i, u_{i-1}, u_{i-2}. The corresponding approximation is known as a one-sided formula. This is sometimes useful to impose boundary conditions on u_x at the ends of the mesh.

3.1.3. Higher-order derivatives

In general, we can derive an approximation of the second derivative using the Taylor expansion

$$\frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x^2} = (\alpha + \beta + \gamma)\frac{1}{\Delta x^2}u_i + (\alpha - \gamma)\frac{1}{\Delta x}\, u_x|_i + (\alpha + \gamma)\frac{1}{2}\, u_{xx}|_i + (\alpha - \gamma)\frac{\Delta x}{6}\, u_{xxx}|_i + (\alpha + \gamma)\frac{\Delta x^2}{12}\, u_{xxxx}|_i + O(\Delta x^4). \tag{25}$$

Using similar arguments to those of the previous section we impose

$$\left.\begin{array}{ll} \alpha+\beta+\gamma & =0 \\ \alpha-\gamma & =0 \\ \alpha+\gamma & =2 \end{array}\right\} \Rightarrow \alpha=\gamma=1,\ \beta=-2. \tag{26}$$

The first and second conditions require that there are no u or u_x terms on the right-hand side of Eq. (25) whilst the third conditon ensures that the right-hand side approximates the left-hand side as Δx tens to zero. The solution of Eq. (26) lead us to the second-order centred approximation

$$u_{xx}|_i = \frac{u_{i+1}-2u_i+u_{i-1}}{\Delta x^2} + O(\Delta x^2). \tag{27}$$

The last term in the Taylor expansion $(\alpha-\gamma)\Delta x u_{xxx}|_i/6$ has the same coefficient as the u_x terms and cancels out to make the approximation second-order accurate. This cancellation does not occur if the points in the mesh are not equally spaced. The derivation of a general three point finite difference approximation with unevenly spaced points can also be obtained through Taylor series. We leave this as an exercise for the reader and proceed in the next section to derive a general form using an alternative method.

3.1.4. Finite differences through polynomial interpolation

In this section we seek to approximate the values of $u(x)$ and its derivatives by a polynomial $P(x)$ at a given point x_i. As way of an example we will derive similar expressions to the centred differences presented previously by considering an approximation involving the set of points $\{x_{i-1}, x_i, x_{i+1}\}$ and the corresponding values $\{u_{i-1}, u_i, u_{i+1}\}$. The polynomial of minimum degree that satisfies $P(x_{i-1}) = u_{i-1}$, $P(x_i) = u_i$ and $P(x_{i+1}) = u_{i+1}$ is the quadratic Lagrange polynomial

$$\begin{aligned} P(x) = u_{i-1}&\frac{(x-x_i)(x-x_{i+1})}{(x_{i-1}-x_i)(x_{i-1}-x_{i+1})} + u_i\frac{(x-x_{i-1})(x-x_{i+1})}{(x_i-x_{i-1})(x_i-x_{i+1})} \\ &+ u_{i+1}\frac{(x-x_{i-1})(x-x_i)}{(x_{i+1}-x_{i-1})(x_{i+1}-x_i)}. \end{aligned} \tag{28}$$

We can now obtain an approximation of the derivative, $u_x|_i \approx P_x(x_i)$ as

$$\begin{aligned} P_x(x_i) = u_{i-1}&\frac{(x_i-x_{i+1})}{(x_{i-1}-x_i)(x_{i-1}-x_{i+1})} + u_i\frac{(x_i-x_{i-1})+(x_i-x_{i+1})}{(x_i-x_{i-1})(x_i-x_{i+1})} \\ &+ u_{i+1}\frac{(x_i-x_{i-1})}{(x_{i+1}-x_{i-1})(x_{i+1}-x_i)}. \end{aligned} \tag{29}$$

If we take $x_i - x_{i-1} = x_{i+1} - x_i = \Delta x$, we recover the second-order accurate finite difference approximation (14) which is consistent with a quadratic

interpolation. Similarly, for the second derivative we have

$$P_{xx}(x_i) = \frac{2u_{i-1}}{(x_{i-1}-x_i)(x_{i-1}-x_{i+1})} + \frac{2u_i}{(x_i-x_{i-1})(x_i-x_{i+1})} + \frac{2u_{i+1}}{(x_{i+1}-x_{i-1})(x_{i+1}-x_i)} \tag{30}$$

and, again, this approximation leads to the second-order centred finite difference (27) for a constant Δx.

This result is general and the approximation via finite differences can be interpreted as a form of Lagrangian polynomial interpolation. The order of the interpolated polynomial is also the order of accuracy of the finite diference approximation using the same set of points. This is also consistent with the interpretation of a Taylor expansion as an interpolating polynomial.

3.1.5. Finite difference solution of PDEs

We consider the FDM approximation to the solution of the elliptic equation $u_{xx} = s(x)$ in the region $\Omega = \{x : 0 \leq x \leq 1\}$. Discretizing the region using N points with constant mesh spacing $\Delta x = (1/N - 1)$ or $x_i = (i - 1/N - 1)$, we consider two cases with different sets of boundary conditions:

1. $u(0) = \alpha_1$ and $u(1) = \alpha_2$, and
2. $u(0) = \alpha_1$ and $u_x(1) = g$.

In both cases we adopt a centred finite approximation in the interior points of the form

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} = s_i; \quad i = 2, \ldots, N-1. \tag{31}$$

The treatment of the first case is straightforward as the boundary conditions are easily specified as $u_1 = \alpha_1$ and $u_N = \alpha_2$. These two conditions together with the $N-2$ equations (31) result in the linear system of N equations with N unknowns represented by

$$\begin{bmatrix} 1 & 0 & & \ldots & & & 0 \\ 1 & -2 & 1 & 0 & \ldots & & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & \ldots & 0 & 1 & -2 & 1 & 0 \\ 0 & & \ldots & 0 & 1 & -2 & 1 \\ 0 & & & \ldots & & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \\ u_N \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \Delta x^2 s_2 \\ \Delta x^2 s_3 \\ \vdots \\ \Delta x^2 s_{N-2} \\ \Delta x^2 s_{N-1} \\ \alpha_2 \end{bmatrix}.$$

This matrix system can be written in abridged form as $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{s}$. The matrix $\boldsymbol{A}$ is non-singular and admits a unique solution $\boldsymbol{u}$. This is the case for most discretization of well-posed elliptic equations.

In the second case the boundary condition $u(0) = \alpha_1$ is treated in the same way by setting $u_1 = \alpha_1$. The treatment of the Neumann boundary condition $u_x(1) = g$ requires a more careful consideration. One possibility is to use a one-sided approximation of $u_x(1)$ to obtain

$$u_x(1) \approx \frac{u_N - u_{N-1}}{\Delta x} = g. \tag{32}$$

This expression is only first-order accurate and thus inconsistent with the approximation used at the interior points. Given that the PDE is elliptic, this error could potentially reduce the global accuracy of the solution. The alternative is to use a second-order centred approximation

$$u_x(1) \approx \frac{u_{N+1} - u_{N-1}}{\Delta x} = g. \tag{33}$$

Here the value u_{N+1} is not available since it is not part of our discrete set of values but we could use the finite difference approximation at x_N given by

$$\frac{u_{N+1} - 2u_N + u_{N-1}}{\Delta x^2} = s_N$$

and include the Neumann boundary condition (33) to obtain

$$u_N - u_{N-1} = \frac{1}{2}(g\,\Delta x - s_N \Delta x^2). \tag{34}$$

It is easy to verify that the introduction of either of the Neumann boundary conditions (32) or (34) leads to non-symmetric matrices.

3.2. Time Integration

In this section we address the problem of solving time-dependent PDEs in which the solution is a function of space and time $u(x, t)$. Consider for instance the heat equation

$$u_t - bu_{xx} = s(x) \text{ in } \Omega = \{x, t : 0 \leq x \leq 1, 0 \leq t \leq T\}$$

with an initial condition $u(x, 0) = u_0(x)$ and time-dependent boundary conditions $u(0, t) = \alpha_1(t)$ and $u(1, t) = \alpha_2(t)$, where α_1 and α_2 are known

functions of t. Assume, as before, a mesh or spatial discretization of the domain $\{x_1, \ldots, x_N\}$.

3.2.1. Method of lines

In this technique we assign to our mesh a set of values that are functions of time $u_i(t) = u(x_i, t)$; $i = 1, \ldots, N$. Applying a centred discretization to the spatial derivative of u leads to a system of ordinary differential equations (ODEs) in the variable t given by

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{b}{x^2}\{u_{i-1}(t) - 2u_i(t) + u_{i+1}(t)\} + s_i; \quad i = 2, \ldots, N-1$$

with $u_1 = \alpha_1(t)$ and $u_N = \alpha_2(t)$. This can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \end{bmatrix} = \frac{b}{\Delta x^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}\begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \end{bmatrix} + \begin{bmatrix} s_2 + \dfrac{b\alpha_1(t)}{\Delta x^2} \\ s_3 \\ \vdots \\ s_{N-2} \\ s_{N-1} + \dfrac{b\alpha_2(t)}{\Delta x^2} \end{bmatrix}$$

or in matrix form as

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}(t) = \mathbf{A}\,\mathbf{u}(t) + \mathbf{s}(t). \tag{35}$$

Equation (35) is referred to as the *semi-discrete* form or the method of lines. This system can be solved by any method for the integration of initial-value problems [3]. The numerical stability of time integration schemes depends on the eigenvalues of the matrix $\mathbf{A}$ which results from the space discretization. For this example, the eigenvalues vary between 0 and $-(4\alpha/\Delta x^2)$ and this could make the system very *stiff*, i.e., with large differences in eigenvalues, as $\Delta x \to 0$.

3.2.2. Finite differences in time

The method of finite differences can be applied to time-dependent problems by considering an independent discretization of the solution $u(x, t)$ in space and time. In addition to the spatial discretization $\{x_1, \ldots, x_N\}$, the discretization in time is represented by a sequence of times $t^0 = 0 < \cdots < t^n < \cdots < T$. For simplicity we will assume constant intervals Δx and Δt in space and time, respectively. The discrete solution at a point will be represented by

$u_i^n \approx u(x_i, t^n)$ and the finite difference approximation of the time derivative follows the procedures previously described. For example, the forward difference in time is given by

$$u_t(x, t^n) \approx \frac{u(x, t^{n+1}) - u(x, t^n)}{\Delta t}$$

and the backward difference in time is

$$u_t(x, t^{n+1}) \approx \frac{u(x, t^{n+1}) - u(x, t^n)}{\Delta t}$$

both of which are first-order accurate, i.e., $\epsilon_{\mathrm{T}} = O(\Delta t)$.

Returning to the heat equation $u_t - bu_{xx} = 0$ and using a centred approximation in space, different schemes can be devised depending on the time at which the equation is discretized. For instance, the use of forward differences in time leads to

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{b}{\Delta x^2}\left(u_{i-1}^n - 2u_i^n + u_{i+1}^n\right). \tag{36}$$

This scheme is *explicit* as the values of the solution at time t^{n+1} are obtained directly from the corresponding (known) values at time t^n. If we use backward differences in time, the resulting scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{b}{\Delta x^2}\left(u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}\right). \tag{37}$$

Here to obtain the values at t^{n+1} we must solve a tri-diagonal system of equations. This type of schemes are referred to as *implicit* schemes.

The higher cost of the implicit schemes is compensated by a greater numerical stability with respect to the explicit schemes which are stable in general only for some combinations of Δx and Δt.

3.3. Discretizations Based on the Integral Form

The FDM uses the strong or differential form of the governing equations. In the following, we introduce two alternative methods that use their integral form counterparts: the finite element and the finite volume methods. The use of integral formulations is advantageous as it provides a more natural treatment of Neumann boundary conditions as well as that of discontinuous source terms due to their reduced requirements on the regularity or smoothness of the solution. Moreover, they are better suited than the FDM to deal with complex geometries in multi-dimensional problems as the integral formulations do not rely in any special mesh structure.

These methods use the integral form of the equation as the starting point of the discretization process. For example, if the strong form of the PDE is $\mathcal{L}(u) = s$, the integral from is given by

$$\int_0^1 \mathcal{L}(u) w(x)\, \mathrm{d}x = \int_0^1 s\, w(x)\, \mathrm{d}x \tag{38}$$

where the choice of the weight function $w(x)$ defines the type of scheme.

3.3.1. *The finite element method (FEM)*

Here we discretize the region of interest $\Omega = \{x : 0 \leq x \leq 1\}$ into $N - 1$ subdomains or elements $\Omega_i = \{x : x_{i-1} \leq x \leq x_i\}$ and assume that the approximate solution is represented by

$$u^\delta(x, t) = \sum_{i=1}^{N} u_i(t) N_i(x)$$

where the set of functions $N_i(x)$ is known as the expansion basis. Its *support* is defined as the set of points where $N_i(x) \neq 0$. If the support of $N_i(x)$ is the whole interval, the method is called a *spectral method*. In the following we will use expansion bases with compact support which are piecewise continuous polynomials within each element as shown in Fig. 6.

The global shape functions $N_i(x)$ can be split within an element into two local contributions of the form shown in Fig. 7. These individual functions are referred to as the *shape functions* or *trial* functions.

3.3.2. *Galerkin FEM*

In the Galerkin FEM method we set the weight function $w(x)$ in Eq. (38) to be the same as the basis function $N_i(x)$, i.e., $w(x) = N_i(x)$.

Consider again the elliptic equation $\mathcal{L}(u) = u_{xx} = s(x)$ in the region Ω with boundary conditions $u(0) = \alpha$ and $u_x(1) = g$. Equation (38) becomes

$$\int_0^1 w(x) u_{xx}\, \mathrm{d}x = \int_0^1 w(x) s(x)\, \mathrm{d}x.$$

At this stage, it is convenient to integrate the left-hand side by parts to get the weak form

$$-\int_0^1 w_x\, u_x\, \mathrm{d}x + w(1)\, u_x(1) - w(0)\, u_x(0) = \int_0^1 w(x)\, s(x)\, \mathrm{d}x. \tag{39}$$

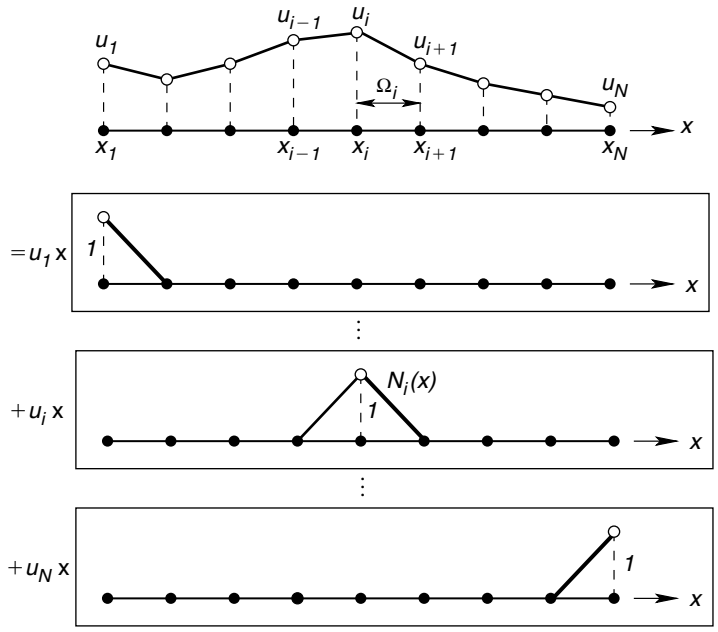

Figure 6. A piecewise linear approximation $u^\delta(x,t) = \sum_{i=1}^{N} u_i(t) N_i(x)$.

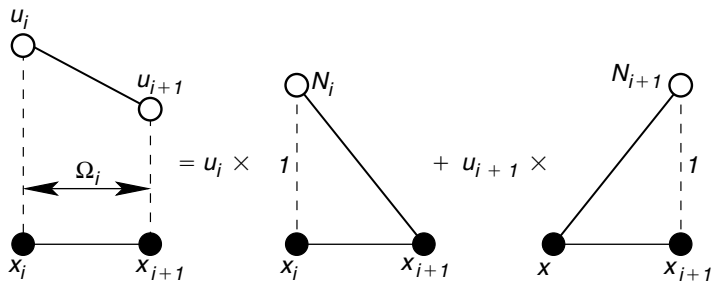

Figure 7. Finite element expansion bases.

This is a common technique in the FEM because it reduces the smoothness requirements on u and it also makes the matrix of the discretized system symmetric. In two and three dimensions we would use Gauss' divergence theorem to obtain a similar result.

The application of the boundary conditions in the FEM deserves attention. The imposition of the Neumann boundary condition $u_x(1) = g$ is straightforward, we simply substitute the value in Eq. (39). This is a very natural way of imposing Neumann boundary conditions which also leads to symmetric

matrices, unlike the FDM. The Dirichlet boundary condition $u(0)=\alpha$ can be applied by imposing $u_1=\alpha$ and requiring that $w(0)=0$. In general, we will impose that the weight functions $w(x)$ are zero at the Dirichlet boundaries.

Letting $u(x)\approx u^\delta(x)=\sum_{j=1}^N u_j N_j(x)$ and $w(x)=N_i(x)$ then Eq. (39) becomes

$$-\int_0^1 \frac{\mathrm{d}N_i}{\mathrm{d}x}(x)\sum_{j=1}^N u_j \frac{\mathrm{d}N_j}{\mathrm{d}x}(x)\,\mathrm{d}x = \int_0^1 N_i(x)\,s(x)\,\mathrm{d}x \tag{40}$$

for $i=2,\ldots,N$. This represents a linear system of $N-1$ equations with $N-1$ unknowns: $\{u_2,\ldots,u_N\}$. Let us proceed to calculate the integral terms corresponding to the ith equation. We calculate the integrals in Eq. (40) as sums of integrals over the elements Ω_i. The basis functions have compact support, as shown in Fig. 6. Their value and their derivatives are different from zero only on the elements containing the node i, i.e.,

$$N_i(x)=\begin{cases}\dfrac{x-x_{i-1}}{\Delta x_{i-1}} & x_{i-1}<x<x_i\\[2ex] \dfrac{x_{i+1}-x}{\Delta x_i} & x_i<x<x_{i+1}\end{cases}$$

$$\frac{\mathrm{d}N_i(x)}{\mathrm{d}x}=\begin{cases}\dfrac{1}{\Delta x_{i-1}} & x_{i-1}<x<x_i\\[2ex] \dfrac{-1}{\Delta x_i} & x_i<x<x_{i+1}\end{cases}$$

with $\Delta x_{i-1}=x_i-x_{i-1}$ and $\Delta x_i=x_{i+1}-x_i$. This means that the only integrals different from zero in (40) are

$$\begin{aligned}&-\int_{x_{i-1}}^{x_i}\frac{\mathrm{d}N_i}{\mathrm{d}x}\left(u_{i-1}\frac{\mathrm{d}N_{i-1}}{\mathrm{d}x}+u_i\frac{\mathrm{d}N_i}{\mathrm{d}x}\right)-\int_{x_i}^{x_{i+1}}\frac{\mathrm{d}N_i}{\mathrm{d}x}\left(u_i\frac{\mathrm{d}N_i}{\mathrm{d}x}+u_{i+1}\frac{\mathrm{d}N_{i+1}}{\mathrm{d}x}\right)\mathrm{d}x\\ &\quad=\int_{x_{i-1}}^{x_i}N_i s\,\mathrm{d}x+\int_{x_i}^{x_{i+1}}N_i s\,\mathrm{d}x\end{aligned} \tag{41}$$

The right-hand side of this equation expressed as

$$F=\int_{x_{i-1}}^{x_i}\frac{x-x_{i-1}}{\Delta x_{i-1}}\,s(x)\,\mathrm{d}x+\int_{x_i}^{x_{i+1}}\frac{x_{i+1}-x}{\Delta x_i}\,s(x)\,\mathrm{d}x$$

can be evaluated using a simple integration rule like the trapezium rule

$$\int_{x_i}^{x_{i+1}} g(x)\,\mathrm{d}x\approx\frac{g(x_i)+g(x_{i+1})}{2}\,\Delta x_i$$

and it becomes

$$F = \left(\frac{\Delta x_{i-1}}{2} + \frac{\Delta x_i}{2} \right) s_i .$$

Performing the required operations in the left-hand side of Eq. (41) and including the calculated valued of F leads to the FEM discrete form of the equation as

$$-\frac{u_i - u_{i-1}}{\Delta x_{i-1}} + \frac{u_{i+1} - u_i}{\Delta x_i} = \frac{\Delta x_{i-1} + \Delta x_i}{2} s_i .$$

Here if we assume that $\Delta x_{i-1} = \Delta x_i = \Delta x$ then the equispaced approximation becomes

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x} = \Delta x \, s_i$$

which is identical to the finite difference formula. We note, however, that the general FE formulation did not require the assumption of an equispaced mesh.

In general the evaluation of the integral terms in this formulation are more efficiently implemented by considering most operations in a standard element $\Omega_{st} = \{-1 \leq x \leq 1\}$ where a mapping is applied from the element Ω_i to the standard element Ω_{st}. For more details on the general formulation see Ref. [4].

3.3.3. Finite volume method (FVM)

The integral form of the one-dimensional linear advection equation is given by Eq. (1) with $f(u) = au$ and $S = 0$. Here the region of integration is taken to be a *control volume* Ω_i, associated with the point of coordinate x_i, represented by $x_{i-(1/2)} \leq x \leq x_{i+(1/2)}$, following the notation of Fig. 4, and the integral form is written as

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} u_t \, \mathrm{d}x + \int_{x_{i-(1/2)}}^{x_{i+(1/2)}} f_x(u) \, \mathrm{d}x = 0. \tag{42}$$

This expression could also been obtained from the weighted residual form (4) by selecting a weight $w(x)$ such that $w(x) = 1$ for $x_{i-(1/2)} \leq x \leq x_{i+(1/2)}$ and $w(x) = 0$ elsewhere. The last term in Eq. (42) can be evaluated analytically to obtain

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} f_x(u) \, \mathrm{d}x = f\left(u_{i+(1/2)}\right) - f\left(u_{i-(1/2)}\right)$$

and if we approximate the first integral using the mid-point rule we finally have the semi-discrete form

$$u_t|_i \left(x_{i+(1/2)} - x_{i-(1/2)}\right) + f\left(u_{i+(1/2)}\right) - f\left(u_{i-(1/2)}\right) = 0.$$

This approach produces a *conservative* scheme if the flux on the boundary of one cell equals the flux on the boundary of the adjacent cell. Conservative schemes are popular for the discretization of hyperbolic equations since, if they converge, they can be proven (Lax-Wendroff theorem) to converge to a weak solution of the conservation law.

3.3.4. Comparison of FVM and FDM

To complete our comparison of the different techniques we consider the FVM discretization of the elliptic equation $u_{xx} = s$. The FVM integral form of this equation over a control volume $\Omega_i = \{x_{i-(1/2)} \leq x \leq x_{i+(1/2)}\}$ is

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} u_{xx}\, \mathrm{d}x = \int_{x_{i-(1/2)}}^{x_{i+(1/2)}} s\, \mathrm{d}x.$$

Evaluating exactly the left-hand side and approximating the right-hand side by the mid-point rule we obtain

$$u_x\left(x_{i+(1/2)}\right) - u_x\left(x_{i-(1/2)}\right) = \left(x_{i+(1/2)} - x_{i-(1/2)}\right) s_i. \tag{43}$$

If we approximate $u(x)$ as a linear function between the mesh points $i-1$ and i, we have

$$u_x|_{i-(1/2)} \approx \frac{u_i - u_{i-1}}{x_i - x_{i-1}}, \qquad u_x|_{i+(1/2)} \approx \frac{u_{i+1} - u_i}{x_{i+1} - x_i},$$

and introducing these approximations into Eq. (43) we now have

$$\frac{u_{i+1} - u_i}{x_{i+1} - x_i} - \frac{u_i - u_{i-1}}{x_i - x_{i-1}} = \left(x_{i+(1/2)} - x_{i-(1/2)}\right) s_i.$$

If the mesh is equispaced then this equation reduces to

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x} = \Delta x\, s_i,$$

which is the same as the FDM and FEM on an equispaced mesh.

Once again we see the similarities that exist between these methods although some assumptions in the construction of the FVM have been made. FEM and FVM allow a more general approach to non-equispaced meshes (although this can also be done in the FDM). In two and three dimensions, curvature is more naturally dealt with in the FVM and FEM due to the integral nature of the equations used.

4. High Order Discretizations: Spectral Element/p-Type Finite Elements

All of the approximations methods we have discussed this far have dealt with what is typically known as the h-type approximation. If $h = \Delta x$ denotes the size of a finite difference spacing or finite elemental regions then convergence of the discrete approximation to the PDE is achieved by letting $h \to 0$. An alternative method is to leave the mesh spacing fixed but to increase the polynomial order of the local approximation which is typically denoted by p or the p-type extension.

We have already seen that higher order finite difference approximations can be derived by fitting polynomials through more grid points. The drawback of this approach is that the finite difference stencil gets larger as the order of the polynomial approximation increases. This can lead to difficulties when enforcing boundary conditions particularly in multiple dimensions. An alternative approach to deriving high order finite differences is to use compact finite differences where a Padé approximation is used to approximate the derivatives.

When using the finite element method in an integral formulation, it is possible to develop a compact high-order discretization by applying higher order polynomial expansions within every elemental region. So instead of using just a linear element in each piecewise approximation of Fig. 6 we can use a polynomial of order p. This technique is commonly known as *p-type finite element* in structural mechanics or the *spectral element* method in fluid mechanics. The choice of the polynomial has a strong influence on the numerical conditioning of the approximation and we note that the choice of an equi-spaced Lagrange polynomial is particularly bad for $p > 5$. The two most commonly used polynomial expansions are Lagrange polynomial based on the Gauss–Lobatto–Legendre quadratures points or the integral of the Legendre polynomials in combination with the linear finite element expansion. These two polynomial expansions are shown in Fig. 8. Although this method is more

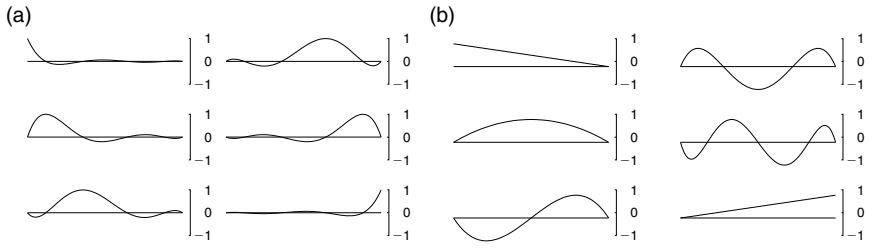

Figure 8. Shape of the fifth order ($p = 5$) polynomial expansions typically used in (a) spectral element and (b) p-type finite element methods.

involved to implement, the advantage is that for a smooth problem (i.e., one where the derivatives of the solution are well behaved) the computational cost increases algebraically whilst the error decreases exponentially fast. Further details on these methods can be found in Refs. [5, 6].

5. Numerical Difficulties

The discretization of linear elliptic equations with either FD, FE or FV methods leads to non-singular systems of equations that can easily solved by standard methods of solution. This is not the case for time-dependent problems where numerical errors may grow unbounded for some discretization. This is perhaps better illustrated with some examples.

Consider the parabolic problem represented by the diffusion equation $u_t - u_{xx} = 0$ with boundary conditions $u(0) = u(1) = 0$ solved using the scheme (36) with $b = 1$ and $\Delta x = 0.1$. The results obtained with $\Delta t = 0.004$ and 0.008 are depicted in Figs. 9(a) and (b), respectively. The numerical solution (b) corresponding to $\Delta t = 0.008$ is clearly unstable.

A similar situation occurs in hyperbolic problems. Consider the one-dimensional linear advection equation $u_t + au_x = 0$; with $a > 0$ and various explicit approximations, for instance the backward in space, or upwind, scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_i^n - u_{i-1}^n}{\Delta x} = 0 \quad \Rightarrow \quad u_i^{n+1} = (1 - \sigma)u_i^n + \sigma u_{i-1}^n, \qquad (44)$$

the forward in space, or downwind, scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_{i+1}^n - u_i^n}{\Delta x} = 0 \quad \Rightarrow \quad u_i^{n+1} = (1 + \sigma)u_i^n - \sigma u_{i+1}^n, \qquad (45)$$

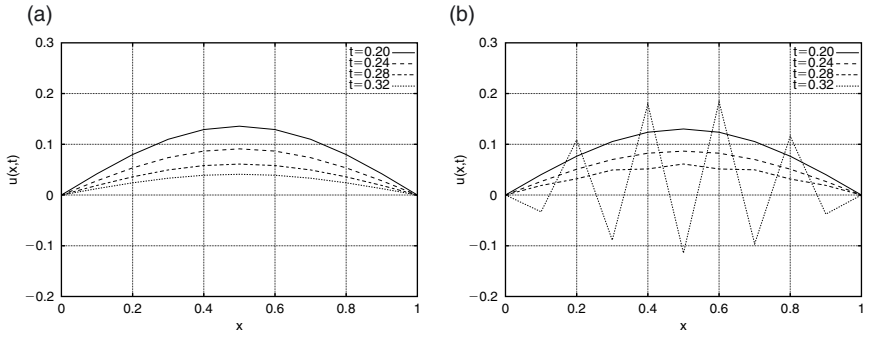

Figure 9. Solution to the diffusion equation $u_t + u_{xx} = 0$ using a forward in time and centred in space finite difference discretization with $\Delta x = 0.1$ and (a) $\Delta t = 0.004$, and (b) $\Delta t = 0.008$. The numerical solution in (b) is clearly unstable.

$$u(x,0) = \begin{cases} 0 & x \le -0.2 \\ 1+5x & -0.2 \le x \le 0 \\ 1-5x & 0 \le x \le 0.2 \\ 0 & x \ge 0.2 \end{cases}$$

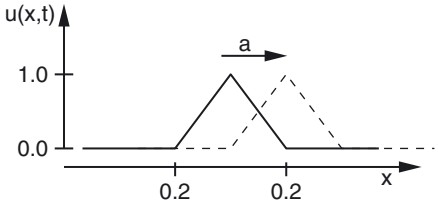

Figure 10. A triangular wave as initial condition for the advection equation.

and, finally, the centred in space is given by

$$\frac{u_i^{n+1}-u_i^n}{\Delta t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0 \quad \Rightarrow \quad u_i^{n+1} = u_i^n - \frac{\sigma}{2}(u_{i+1}^n - u_{i-1}^n) \tag{46}$$

where $\sigma = (a\Delta t/\Delta x)$ is known as the *Courant number*. We will see later that this number plays an important role in the stability of hyperbolic equations. Let us obtain the solution of $u_t + au_x = 0$ for all these schemes with the initial condition given in Fig. 10.

As also indicated in Fig. 10, the exact solution is the propagation of this wave form to the right at a velocity a. Now we consider the solution of the three schemes at two different Courant numbers given by $\sigma = 0.5$ and 1.5. The results are presented in Fig. 11 and we observe that only the upwinded scheme when $\sigma \le 1$ gives a stable, although diffusive, solution. The centred scheme when $\sigma = 0.5$ appears almost stable but the oscillations grow in time leading to an unstable solution.

6. Analysis of Numerical Schemes

We have seen that different parameters, such as the Courant number, can effect the stability of a numerical scheme. We would now like to set up a more rigorous framework to analyse a numerical scheme and we introduce the concepts of *consistency*, *stability* and *Convergence* of a numerical scheme.

6.1. Consistency

A numerical scheme is consistent if the discrete numerical equation tends to the exact differential equation as the mesh size (represented by Δx and Δt) tends to zero.

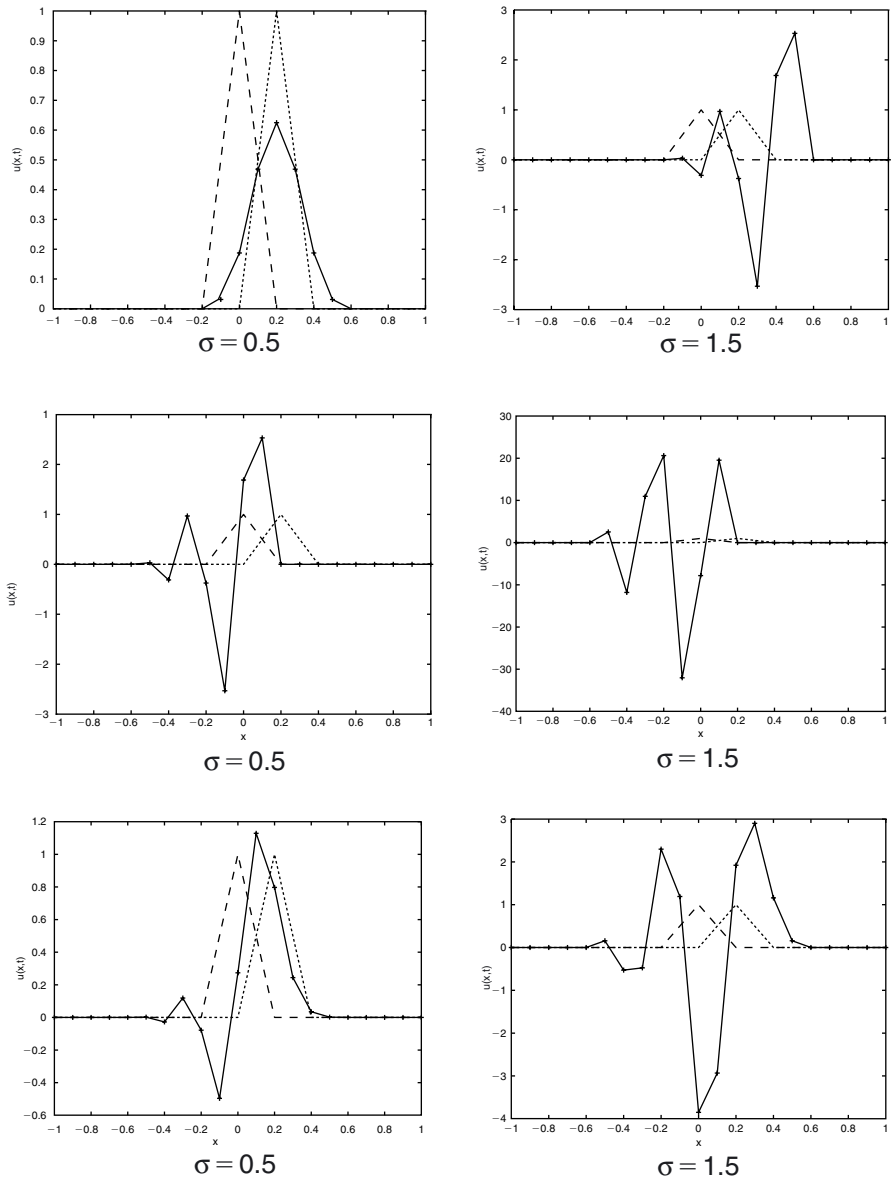

Figure 11. Numerical solution of the advection equation $u_t + au_x = 0$. Dashed lines: initial condition. Dotted lines: exact solution. Solid line: numerical solution.

Consider the centred in space and forward in time finite diference approximation to the linear advection equation $u_t + au_x = 0$ given by Eq. (46). Let us consider Taylor expansions of u_i^{n+1}, u_{i+1}^n and u_{i-1}^n around (x_i, t^n) as

$$u_i^{n+1} = u_i^n + \Delta t\ u_t|_i^n + \frac{\Delta t^2}{2} u_{tt}|_i^n + \cdots$$

$$u_{i+1}^n = u_i^n + \Delta x\, u_x|_i^n + \frac{\Delta x^2}{2} u_{xx}|_i^n + \frac{\Delta x^3}{6} u_{xxx}|_i^n + \cdots$$
$$u_{i-1}^n = u_i^n - \Delta x\, u_x|_i^n + \frac{\Delta x^2}{2} u_{xx}|_i^n - \frac{\Delta x^3}{6} u_{xxx}|_i^n + \cdots$$

Substituting these expansions into Eq. (46) and suitably re-arranging the terms we find that

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} - (u_t + a u_x)|_i^n = \epsilon_{\mathrm{T}} \tag{47}$$

where ϵ_{T} is known as the *truncation error* of the approximation and is given by

$$\epsilon_{\mathrm{T}} = \frac{\Delta t}{2} u_{tt}|_i^n + \frac{\Delta x^2}{6} a u_{xxx}|_i^n + O(\Delta t^2, \Delta x^4).$$

The left-hand side of this equation will tend to zero as Δt and Δx tend to zero. This means that the numerical scheme (46) tends to the exact equation at point x_i and time level t^n and therefore this approximation is *consistent*.

6.2. Stability

We have seen in the previous numerical examples that errors in numerical solutions can grow uncontrolled and render the solution meaningless. It is therefore sensible to require that the solution is stable, this is that the difference between the computed solution and the exact solution of the discrete equation should remain bounded as $n \to \infty$ for a given Δx.

6.2.1. The Courant–Friedrichs–Lewy (CFL) condition

This is a necessary condition for stability of explicit schemes devised by Courant, Friedrichs and Lewy in 1928.

Recalling the theory of characteristics for hyperbolic systems, the *domain of dependence of a PDE* is the portion of the domain that influences the solution at a given point. For a scalar conservation law, it is the characteristic passing through the point, for instance, the line PQ in Fig. 12. The *domain of dependence of a FD scheme* is the set of points that affect the approximate solution at a given point. For the upwind scheme, the numerical domain of dependence is shown as a shaded region in Fig. 12.

The *CFL criterion* states that a *necessary* condition for an explicit FD scheme to solve a hyperbolic PDE to be stable is that, for each mesh point, the domain of dependence of the FD approximation contains the domain of dependence of the PDE.

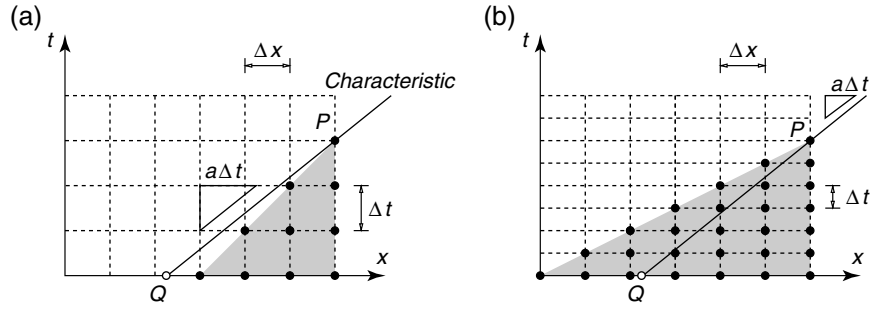

Figure 12. Solution of the advection equation by the upwind scheme. Physical and numerical domains of dependence: (a) $\sigma = (a\Delta t/\Delta x) > 1$, (b) $\sigma \leq 1$.

For a Courant number $\sigma = (a\Delta t/\Delta x)$ greater than 1, changes at Q will affect values at P but the FD approximation cannot account for this.

The CFL condition is necessary for stability of explicit schemes but *it is not sufficient*. For instance, in the previous schemes we have that the upwind FD scheme is stable if the CFL condition $\sigma \leq 1$ is imposed. The downwind FD scheme does not satisfy the CFL condition and is unstable. However, the centred FD scheme is unstable even if $\sigma \leq 1$.

6.2.2. *Von Neumann (or Fourier) analysis of stability*

The stability of FD schemes for hyperbolic and parabolic PDEs can be analysed by the von Neumann or Fourier method. The idea behind the method is the following. As discussed previously the analytical solutions of the model diffusion equation $u_t - b\,u_{xx} = 0$ can be found in the form

$$u(x,t) = \sum_{m=-\infty}^{\infty} \mathrm{e}^{\beta_m t}\,\mathrm{e}^{I k_m x}$$

if $\beta_m + b\,k_m^2 = 0$. This solution involves a Fourier series in space and an exponential decay in time since $\beta_m \leq 0$ for $b > 0$. Here we have included the complex version of the Fourier series, $\mathrm{e}^{I k_m x} = \cos\, k_m x + I\, \sin\, k_m x$ with $I = \sqrt{-1}$, because this simplifies considerably later algebraic manipulations.

To analyze the growth of different Fourier modes as they evolve under the numerical scheme we can consider each frequency separately, namely

$$u(x,t) = \mathrm{e}^{\beta_m t}\,\mathrm{e}^{I k_m x}.$$

A discrete version of this equation is $u_i^n = u(x_i, t^n) = \mathrm{e}^{\beta_m t^n}\mathrm{e}^{Ik_m x_i}$. We can take, without loss of generality, $x_i = i\,\Delta x$ and $t^n = n\,\Delta t$ to obtain

$$u_i^n = \mathrm{e}^{\beta_m n\,\Delta t}\,\mathrm{e}^{Ik_m i\,\Delta x} = \left(\mathrm{e}^{\beta_m\,\Delta t}\right)^n \mathrm{e}^{Ik_m i\,\Delta x}.$$

The term $\mathrm{e}^{Ik_m i\,\Delta x} = \cos(k_m i\,\Delta x) + I\,\sin(k_m i\,\Delta x)$ is bounded and, therefore, any growth in the numerical solution will arise from the term $G = \mathrm{e}^{\beta_m\,\Delta t}$, known as the *amplification factor*. Therefore the numerical method will be stable, or the numerical solution u_i^n bounded as $n \to \infty$, if $|G| \leq 1$ for solutions of the form

$$u_i^n = G^n\,\mathrm{e}^{Ik_m i\,\Delta x}.$$

We will now proceed to analyse, using the von Neummann method, the stability of some of the schemes discussed in the previous sections.

Example 1 Consider the explicit scheme (36) for the diffusion equation $u_t - bu_{xx} = 0$ expressed here as

$$u_i^{n+1} = \lambda u_{i-1}^n + (1 - 2\lambda)u_i^n + \lambda u_{i+1}^n; \quad \lambda = \frac{b\,\Delta t}{\Delta x^2}.$$

We assume $u_i^n = G^n\mathrm{e}^{Ik_m i\,\Delta x}$ and substitute in the equation to get

$$G = 1 + 2\lambda\,[\cos(k_m\,\Delta x) - 1]\,.$$

Stability requires $|G| \leq 1$. Using $-2 \leq \cos(k_m\,\Delta x) - 1 \leq 0$ we get $1 - 4\lambda \leq G \leq 1$ and to satisfy the left inequality we impose

$$-1 \leq 1 - 4\lambda \leq G \quad \Longrightarrow \quad \lambda \leq \frac{1}{2}.$$

This means that for a given grid size Δx the maximum allowable timestep is $\Delta t = (\Delta x^2/2b)$.

Example 2 Consider the implicit scheme (37) for the diffusion equation $u_t - bu_{xx} = 0$ expressed here as

$$\lambda u_{i-1}^{n+1} + -(1 + 2\lambda)u_i^{n+1} + \lambda u_{i+1}^{n+1} = -u_i^n\,; \quad \lambda = \frac{b\,\Delta t}{\Delta x^2}.$$

The amplification factor is now

$$G = \frac{1}{1 + \lambda(2 - \cos\,\beta_m)}$$

and we have $|G| < 1$ for any β_m if $\lambda > 0$. This scheme is therefore unconditionally stable for any Δx and Δt. This is obtained at the expense of solving a linear system of equations. However, there will still be restrictions on Δx

and Δt based on the accuracy of the solution. The choice between an explicit or an implicit method is not always obvious and should be done based on the computer cost for achieving the required accuracy in a given problem.

Example 3 Consider the upwind scheme for the linear advection equation $u_t + a u_x = 0$ with $a > 0$ given by

$$u_i^{n+1} = (1 - \sigma) u_i^n + \sigma u_{i-1}^n; \quad \sigma = \frac{a \Delta t}{\Delta x}.$$

Let us denote $\beta_m = k_m \Delta x$ and introduce the discrete Fourier expression in the upwind scheme to obtain

$$G = (1 - \sigma) + \sigma \, \mathrm{e}^{-I \beta_m}$$

The stability condition requires $|G| \leq 1$. Recall that G is a *complex* number $G = \xi + I \eta$ so

$$\xi = 1 - \sigma + \sigma \cos \beta_m; \quad \eta = -\sigma \sin \beta_m$$

This represents a circle of radius σ centred at $1 - \sigma$. The stability condition requires the locus of the points (ξ, η) to be interior to a unit circle $\xi^2 + \eta^2 \leq 1$. If $\sigma < 0$ the origin is outside the unit circle, $1 - \sigma > 1$, and the scheme is unstable. If $\sigma > 1$ the back of the locus is outside the unit circle $1 - 2\sigma < 1$ and it is also unstable. Therefore, for stability we require $0 \leq \sigma \leq 1$, see Fig. 13.

Example 4 The forward in time, centred in space scheme for the advection equation is given by

$$u_i^{n+1} = u_i^n - \frac{\sigma}{2}(u_{i+1}^n - u_{i-1}^n); \quad \sigma = \frac{a \Delta t}{\Delta x}.$$

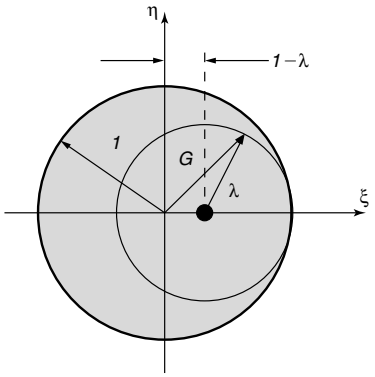

Figure 13. Stability region of the upwind scheme.

The introduction of the discrete Fourier solution leads to

$$G = 1 - \frac{\sigma}{2}(\mathrm{e}^{I\beta_m} - \mathrm{e}^{-I\beta_m}) = 1 - I\sigma \ \sin \ \beta_m$$

Here we have $|G|^2 = 1 + \sigma^2 \ \sin^2 \ \beta_m > 1$ always for $\sigma \neq 0$ and it is therefore unstable. We will require a different time integration scheme to make it stable.

6.3. Convergence: Lax Equivalence Theorem

A scheme is said to be *convergent* if the difference between the computed solution and the exact solution of the PDE, i.e., the error $E_i^n = u_i^n - u(x_i, t^n)$, vanishes as the mesh size is decreased. This is written as

$$\lim_{\Delta x, \Delta t \to 0} |E_i^n| = 0$$

for fixed values of x_i and t^n. This is the fundamental property to be sought from a numerical scheme but it is difficult to verify directly. On the other hand, consistency and stability are easily checked as shown in the previous sections.

The main result that permits the assessment of the convergence of a scheme from the requirements of consistency and stability is the equivalence theorem of Lax stated here without proof:

Stability is the necessary and sufficient condition for a *consistent* linear FD approximation to a well-posed linear initial-value problem to be *convergent*.

7. Suggestions for Further Reading

The basics of the FDM are presented a very accessible form in Ref. [7]. More modern references are Refs. [8, 9].

An elementary introduction to the FVM can be consulted in the book by Versteeg and Malalasekera [10]. An in-depth treatment of the topic with an emphasis on hyperbolic problems can be found in the book by Leveque [2].

Two well established general references for the FEM are the books of Hughes [4] and Zienkiewicz and Taylor [11]. A presentation from the point of view of structural analysis can be consulted in Cook *et al*. [11]

The application of p-type finite element for structural mechanics is dealt with in book of Szabo and Babuška [5]. The treatment of both p-type and spectral element methods in fluid mechanics can be found in book by Karniadakis and Sherwin [6].

A comprehensive reference covering both FDM, FVM and FEM for fluid dynamics is the book by Hirsch [13]. These topics are also presented using a more mathematical perspective in the classical book by Quarteroni and Valli [14].

References

[1] J. Bonet and R. Wood, *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press, 1997.

[2] R. Leveque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, 2002.

[3] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing*, 4th edn., Brooks/Cole Publishing Co., 1999.

[4] T. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Dover Publishers, 2000.

[5] B. Szabo and I. Babuška, *Finite Element Analysis*, Wiley, 1991.

[6] G.E. Karniadakis and S. Sherwin, *Spectral/hp Element Methods for CFD*, Oxford University Press, 1999.

[7] G. Smith, *Numerical Solution of Partial Differential Equations: Finite Diference Methods*, Oxford University Press, 1985.

[8] K. Morton and D. Mayers, *Numerical Solution of Partial Differential Equations*, Cambridge University Press, 1994.

[9] J. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer-Verlag, 1995.

[10] H. Versteeg and W. Malalasekera, *An Introduction to Computational Fluid Dynamics. The Finite Volume Method*, Longman Scientific & Technical, 1995.

[11] O. Zienkiewicz and R. Taylor, *The Finite Element Method: The Basis*, vol. 1, Butterworth and Heinemann, 2000.

[12] R. Cook, D. Malkus, and M. Plesha, *Concepts and Applications of Finite Element Analysis*, Wiley, 2001.

[13] C. Hirsch, *Numerical Computation of Internal and External Flows*, vol. 1, Wiley, 1988.

[14] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, 1994.