

$$\begin{aligned}
 (2.8-3) \quad & u(ih, (n+1)k) - u(ih, (n-1)k) \\
 & - \frac{2k}{h^2} [u((i+1)h, k) - u(ih, (n+1)k) \\
 & - u(ih, (n-1)k) + u((i-1)h, k)] \\
 & = -2k \left[\left(\frac{k}{h} \right)^2 \frac{\partial^2 u}{\partial t^2} + \frac{k^2}{6} \frac{\partial^3 u}{\partial t^3} - \frac{h^2}{12} \frac{\partial^4 u}{\partial y^4} + \dots \right]_{(ih, nk)}.
 \end{aligned}$$

This difference formula is referred to as the **DuFort-Frankel approximation**. Consider the asymptotic behavior of the error on the right side of Equation (2.8-3) in the limit as h and k approach zero. Whereas in the previous example the truncation error described by Equation (2.8-2) vanished in the limit as h and k approached zero, irrespective of the relative behavior of h and k , such is not the case with the DuFort-Frankel approximation. If $k/h \rightarrow 0$ as $k \rightarrow 0$, then the error term in Equation (2.8-3) vanishes, and the difference formula approximates the model equation. However, if k/h tends to a nonzero constant as $k \rightarrow 0$, then the difference formula is an approximation to

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} + \left[\lim_{k,h \rightarrow 0} \left(\frac{k}{h} \right)^2 \right] \frac{\partial^2 u}{\partial t^2},$$

which is hyperbolic rather than parabolic. Thus one concludes that the difference expression appearing in Equation (2.8-3) is consistent with the model equation if and only if k goes to zero faster than h , and therefore the DuFort-Frankel approximation is **conditionally consistent** with the heat equation.

2.9. Stability of Finite-Difference Approximations.

The notion of **stability** addresses the computational behavior of the algebraic equations arising from the numerical approximation of differential equations. There are several ways of approaching the concept of stability. Here we define a **stable algorithm** to be one for which every component of an initial function, possibly containing numerical errors, is limited in the degree to which it is amplified by the numerical procedure. Unstable algorithms are virtually useless in digital calculations, since they tend to amplify unavoidable roundoff errors without bound. Let us now consider the question of how to establish the stability of a finite-difference equation.

One approach consists of programming the algorithm for the digital computer and then conducting a series of numerical experiments. Through examination of the behavior of the method over a broad spectrum of mesh geometries and coefficient values, one can come to certain conclusions re-

garding the stability bounds of the method. The approach is often referred to as **heuristic stability analysis**.

A more structured approach was proposed by von Neumann (Charney et al., 1950). It furnishes a necessary and, in the case of certain pure initial value problems, sufficient condition for stability. The procedure is based on Fourier analysis. A reasonably well behaved function $u(x, t)$ on a spatial interval $[0, X]$ can be represented by an infinite series

$$(2.9-1) \quad u(x, t) = \sum_{m=-\infty}^{\infty} g_m(t) e^{i2\pi m(x/X)},$$

where $i = \sqrt{-1}$ and

$$(2.9-2) \quad g_m(t) = \frac{1}{X} \int_0^X u(x, t) e^{-i2\pi m(x/X)} dx.$$

The quantity g_m is the amplitude of the m -th harmonic or Fourier mode; this mode has a wavelength X/m .

Assume we know $u(x, t)$ at $+1L$ distinct points or nodes (ih, t) , $i = 0, 1, 2, \dots, L$. Then the values $u(ih, t)$, $i = 0, 1, 2, \dots, L$, can be approximated by a set of L Fourier modes, where x is replaced by ih and X by Lh :

$$(2.9-3) \quad u(ih, t) \simeq \sum_{m=0}^{L-1} \hat{g}_m(t) e^{i2\pi m(i/L)}.$$

Here the amplitude of the m -th mode is given by a discrete version of the integral (2.9-2):

$$(2.9-4) \quad \hat{g}_m(t) = \frac{1}{L} \sum_{i=0}^{L-1} u(ih, t) e^{-i2\pi m(i/L)}.$$

Note that this representation constitutes a "long wavelength" approximation to $u(x, t)$, which is appropriate since no wavelength smaller than h can be represented on the finite-difference grid.

Our task now is to examine the propagation of errors in finite-difference expressions as the simulation advances through time. Consider once again a difference approximation to the one-dimensional heat equation. Employing a forward-in-time, centered-in-space approximation, we obtain the following algebraic equations for the approximate solution $\hat{u}(x, t)$:

$$(2.9-5) \quad \begin{aligned} & \hat{u}(ih, (n+1)k) - \hat{u}(ih, nk) \\ & - \frac{k}{h^2} [\hat{u}((i+1)h, nk) - 2\hat{u}(ih, nk) + \hat{u}((i-1)h, nk)] = 0. \end{aligned}$$

Let us assume that, at $t = 0$, machine limitations introduce an error at the spatial nodes. We denote this error as $\epsilon(ih, 0)$. The machine will force the function $\hat{u} + \epsilon$ to satisfy the difference equation, while in principle we want \hat{u} to satisfy it. Moreover, let us represent the error ϵ by a Fourier expansion of the form presented in (2.9-3):

$$(2.9-6) \quad \epsilon(ih, 0) = \sum_{m=0}^{L-1} \hat{g}_m(0) e^{i2\pi m(i/L)},$$

where

$$(2.9-7) \quad \hat{g}_m(0) = \frac{1}{L} \sum_{i=0}^{L-1} \epsilon(ih, 0) e^{-i2\pi m(i/L)}.$$

Because, by definition, the true values of \hat{u} satisfy the difference equation, which is linear, the error also satisfies the difference equation. We can thus write

$$(2.9-8) \quad \begin{aligned} & \epsilon(ih, (n+1)k) - \epsilon(ih, nk) \\ &= \frac{k}{h^2} [\epsilon((i+1)h, nk) - 2\epsilon(ih, nk) + \epsilon((i-1)h, nk)]. \end{aligned}$$

From the principle of superposition we know that we can establish the behavior of the scheme (2.9-8) by examining the propagation of a typical term in the series (2.9-6) for ϵ . Such a term has the form

$$(2.9-9) \quad \epsilon_m(ih, nk) = \hat{g}_m(nk) e^{i\beta_m ih},$$

where $\beta_m = 2\pi m/Lh$. To simplify notation we define $\xi^n = \hat{g}_m(nk)$ and designate ξ the amplification factor. Substitution of the term (2.9-9) into the difference equation (2.9-8), suppression of the mode number m , and replacement of k/h^2 by the symbol $\bar{\rho}$ yield

$$(2.9-10) \quad \begin{aligned} & \xi^{n+1} e^{i\beta_i h} - \xi^n e^{i\beta_i h} \\ &= \bar{\rho} [\xi^n e^{i\beta(i+1)h} - 2\xi^n e^{i\beta ih} + \xi^n e^{i\beta(i-1)h}]. \end{aligned}$$

This expression can be solved for the amplification factor, giving

$$(2.9-11) \quad \begin{aligned} \xi &= 1 + \bar{\rho}(e^{i\beta h} - 2 + e^{-i\beta h}) \\ &= (1 - 2\bar{\rho}) + \bar{\rho}(e^{i\beta h} + e^{-i\beta h}) \\ &= 1 - 2\bar{\rho}(1 - \cos \beta h) \\ &= 1 - 4\bar{\rho} \sin^2 \left(\frac{\beta h}{2} \right). \end{aligned}$$

To interpret ξ , observe from Equation (2.9-9) that

$$(2.9-12) \quad \frac{\epsilon(ih, (n+1)k)}{\epsilon(ih, nk)} = \frac{\xi^{n+1}}{\xi^n}$$

or, upon rearranging,

$$(2.9-13) \quad \epsilon(ih, (n+1)k) = \xi \epsilon(ih, nk).$$

Thus the error component ϵ will not grow provided $|\xi| \leq 1$. This is the von Neumann necessary condition for stability.

Returning to Equation (2.9-11) we observe that von Neumann stability for this example requires

$$(2.9-14) \quad \left| 1 - 4\bar{\rho} \sin^2\left(\frac{\beta h}{2}\right) \right| \leq 1.$$

For this inequality to hold for all values of β , it is necessary that

$$(2.9-15) \quad \bar{\rho} \leq \frac{1}{2}$$

or, recalling the definition of $\bar{\rho}$,

$$(2.9-16) \quad k \leq \frac{1}{2}h^2.$$

Thus for a given spatial increment h the scheme (2.9-8) will be stable only for small enough time steps k . Similar relationships can be developed for other difference approximations, albeit with possibly more tedious algebraic manipulations.

A third approach to stability, and one particularly suited to approximation methods exclusive of finite differences, is **matrix stability analysis**. We illustrate the methodology using the same example presented in the preceding section on von Neumann stability.

The finite-difference expression (2.9-8), in combination with Dirichlet boundary conditions, can be written in matrix form as

$$(2.9-17) \quad \mathbf{A}\boldsymbol{\epsilon}^{(n)} + \mathbf{B}(\boldsymbol{\epsilon}^{(n+1)} - \boldsymbol{\epsilon}^{(n)}) = \mathbf{0},$$

where the nonzero elements of \mathbf{A} are given by

$$a_{i,i-1} = -\bar{\rho}, \quad i \neq 1,$$

$$a_{i,i} = 2\bar{\rho},$$

$$a_{i,i+1} = -\bar{\rho}, \quad i \neq L.$$

The matrix \mathbf{B} is diagonal, having nonzero entries $b_{i,i} = 1$, while the error vectors $\boldsymbol{\epsilon}^{(n)}$ and $\boldsymbol{\epsilon}^{(n+1)}$ have entries

$$\begin{aligned}\epsilon_i^{(n)} &= \epsilon(ih, nk), \\ \epsilon_i^{(n+1)} &= \epsilon(ih, (n+1)k).\end{aligned}$$

Equation (2.9-17) can be rearranged to yield $\mathbf{B}\epsilon^{(n+1)} = \mathbf{B}\epsilon^{(n)} - \mathbf{A}\epsilon^{(n)} \equiv \mathbf{C}\epsilon^{(n)}$, where $\mathbf{C} = \mathbf{B} - \mathbf{A}$. Because \mathbf{B} is the identity matrix \mathbf{I} in our example, this expression can be written

$$\epsilon^{(n+1)} = \mathbf{B}^{-1}\mathbf{C}\epsilon^{(n)} = \mathbf{C}\epsilon^{(n)} = \mathbf{C}^n\epsilon^{(0)},$$

where the nonzero elements of \mathbf{C} are as follows

$$\begin{aligned}c_{i,i-1} &= \bar{\rho}, & i &\neq 1, \\ c_{i,i} &= 1 - 2\bar{\rho}, \\ c_{i,i+1} &= \bar{\rho}, & i &\neq L.\end{aligned}$$

The error growth will be bounded provided the norm of the error vector is bounded with respect to the initial error as time progresses. Thus, for any norm $\|\cdot\|$, stability requires

$$(2.9-18) \quad \|\epsilon^{(n+1)}\| \leq K\|\epsilon^{(0)}\| \quad \text{as } n \rightarrow \infty,$$

for some constant K independent of the discretization intervals h and k . Employing the Schwarz inequality and assuming \mathbf{C} is independent of time, we get

$$\begin{aligned}(2.9-19) \quad \|\epsilon^{(n+1)}\| &= \|\mathbf{C}\epsilon^{(n)}\| = \|\mathbf{C}^n\epsilon^{(0)}\| \leq \|\mathbf{C}^n\| \|\epsilon^{(0)}\| \\ &\leq \|\mathbf{C}\|^n \|\epsilon^{(0)}\|.\end{aligned}$$

Here the norm of the matrix \mathbf{C} is defined relative to the vector norm $\|\cdot\|$ via the equation

$$\|\mathbf{C}\| = \sup \frac{\|\mathbf{Cx}\|}{\|\mathbf{x}\|},$$

where "sup" denotes the least upper bound taken over all nonzero vectors \mathbf{x} . Thus a sufficient condition for stability is as follows:

$$(2.9-20) \quad \|\mathbf{C}\|^n \leq K \quad \text{as } n \rightarrow \infty.$$

A more useful relationship can be derived using the relationship between the Euclidean matrix norm and the spectral radius or maximum-modulus eigenvalue, $\rho(\mathbf{C}) = \max_i |\lambda_i(\mathbf{C})|$:

$$(2.9-21) \quad \rho^n(\mathbf{C}) \leq \|\mathbf{C}\|_2^n.$$

Therefore a necessary condition for (2.9-20) to hold is

$$(2.9-22) \quad \rho^n(\mathbf{C}) \leq K \quad \text{as } n \rightarrow \infty,$$

which implies the requirement

$$(2.9-23) \quad \rho(\mathbf{C}) \leq 1.$$

In general the Euclidean norm is related to the spectral radius through the equation

$$(2.9-24) \quad \|\mathbf{C}\|_2 = \sqrt{\rho(\mathbf{C}\mathbf{C}^\top)}.$$

Thus for symmetric matrices we can replace the inequality (2.9-21) by

$$(2.9-25) \quad \|\mathbf{C}\|_2 = \rho(\mathbf{C}),$$

and (2.9-23) becomes both necessary and sufficient.

Let us now return to our example problem for the heat equation. Because of the particular structure of \mathbf{C} , the eigenvalues can be computed analytically in this problem. They are as follows:

$$(2.9-26) \quad \lambda_i(\mathbf{C}) = 1 - 4\bar{\rho} \sin^2\left(\frac{i\pi}{2L}\right), \quad i = 1, 2, \dots, L-1.$$

Hence the criterion (2.9-23) becomes

$$(2.9-27) \quad \left|1 - 4\bar{\rho} \sin^2\left(\frac{i\pi}{2L}\right)\right| \leq 1, \quad i = 1, 2, \dots, L-1,$$

which implies the same stability criterion derived using the von Neumann method.

2.10. The Method of Weighted Residuals.

The **method of weighted residuals** provides a conceptual foundation upon which to construct Galerkin finite-element, collocation, and boundary element methods, all to be considered in this chapter. While we shall become more specific later, at this point let us consider a differential equation written in operator form as

$$(2.10-1) \quad \mathcal{L}(u(\mathbf{x})) - f(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega,$$

where $\mathcal{L}(\cdot)$ is the differential operator and $f(\mathbf{x})$ is a known **forcing function**, both defined over a region Ω with boundary $\partial\Omega$. Let $\phi_i(\mathbf{x})$, $i = 1, 2, \dots, N$ represent N functions selected from a set of known, linearly independent **basis functions**. In general these functions will satisfy homogeneous boundary conditions on $\partial\Omega$, but to see precisely what these conditions should be we need to carry the development a little further. Let